

Article

Short-Term Electricity Demand Forecasting: A Comparative Evaluation of Models Based on Performance Criteria and Future Research Directions

Anderson Sebastian Torres-Sánchez , Álvaro Jaramillo-Duque  and Walter M. Villa-Acevedo * 

Grupo de Investigación en Manejo Eficiente de la Energía (GIMEL), Departamento de Ingeniería Eléctrica, Facultad de Ingeniería, Universidad de Antioquia (UdeA), Calle 67 #54-108, Medellín 050010, Colombia; anderson.torress@udea.edu.co (A.S.T.-S.); alvaro.jaramillod@udea.edu.co (Á.J.-D.)

* Correspondence: walter.villa@udea.edu.co

Abstract

Short-term electricity demand forecasting is a critical enabler of the secure and efficient operation of modern power systems, particularly amid increasing renewable energy integration, smart grid expansion, and the broader energy transition. This paper presents a rigorous comparative analysis of electricity demand forecasting models, encompassing statistical methods, *Machine Learning* (ML), *Deep Learning* (DL), and hybrid architectures. A structured taxonomy is proposed to classify models according to their methodological family, application horizon, and data requirements, thereby providing a unified reference framework for researchers and energy-sector practitioners. Models are evaluated using a multi-criteria framework comprising accuracy, robustness, scalability, interpretability, computational cost, and the capacity to handle exogenous variables. The analysis identifies critical research gaps, including the limited integration of probabilistic forecasting into operational contexts and the absence of standardized evaluation protocols under real-world conditions. Future research directions are outlined, with particular emphasis on uncertainty quantification, adaptive learning strategies, and hierarchical forecast coherence in systems with high penetration of distributed energy resources.

Keywords: short-term load forecasting; electricity demand forecasting; machine learning; deep learning; mixture of experts; statistical models; taxonomy

1. Introduction

Electricity demand forecasting is a fundamental process for the efficient planning, operation, and management of modern power systems. In the current context of increasing electrification, digitalization, and energy transition, the ability to accurately anticipate future demand behavior has become essential to ensure supply security, minimize operating costs, and promote efficient and sustainable management of energy resources [1,2]. This challenge is further intensified by global trends that push traditional electrical infrastructures to their limits, making robust analytical tools essential for reliable and effective decision-making.

In this context, electricity demand is defined as the amount of energy required by system consumers at a given time [3,4]. The sustained growth in global electricity consumption over the last three decades, based on data compiled by [5] and presented in Figure 1, illustrates the profound transformation experienced by power systems worldwide. From nearly 10,000 TWh in 1990 to more than 26,000 TWh in 2024, this evolution reflects not only a substantial increase in electricity demand, but also growing pressure on



Academic Editor: Erol Kurt

Received: 26 May 2026

Revised: 30 June 2026

Accepted: 9 July 2026

Published: 11 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

planning and operational mechanisms. Since consumption directly reflects the aggregate behavior of demand, its sustained increase underscores the need for more accurate, dynamic forecasting models tailored to specific contexts and conditions. This global trend is also observed, with variations, in regional contexts such as Latin America (see Figure 2), where consumption has nearly doubled over the same period. However, this pattern is less pronounced compared to Asia or North America, as illustrated in Figure 3, which was also constructed using data from [5], suggesting the presence of structural differences and highlighting the need to develop forecasting models adapted to the particular dynamics and consumption patterns of each region.

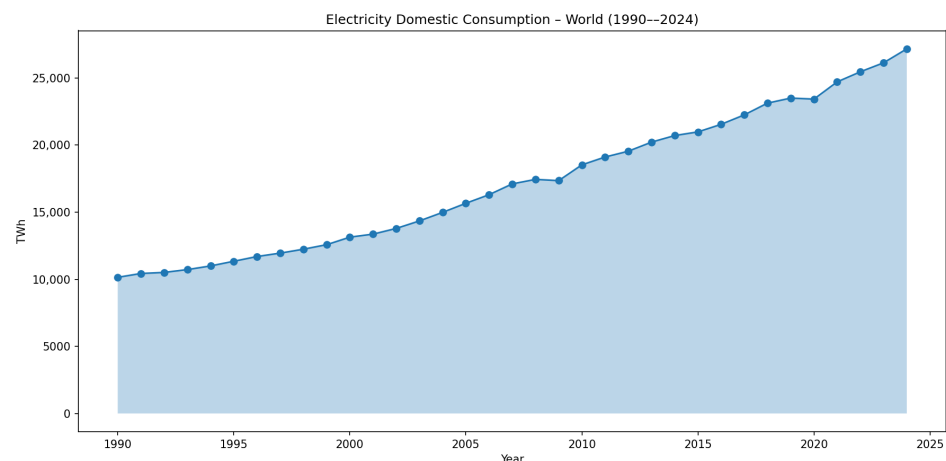


Figure 1. Global electricity consumption between 1990 and 2024.

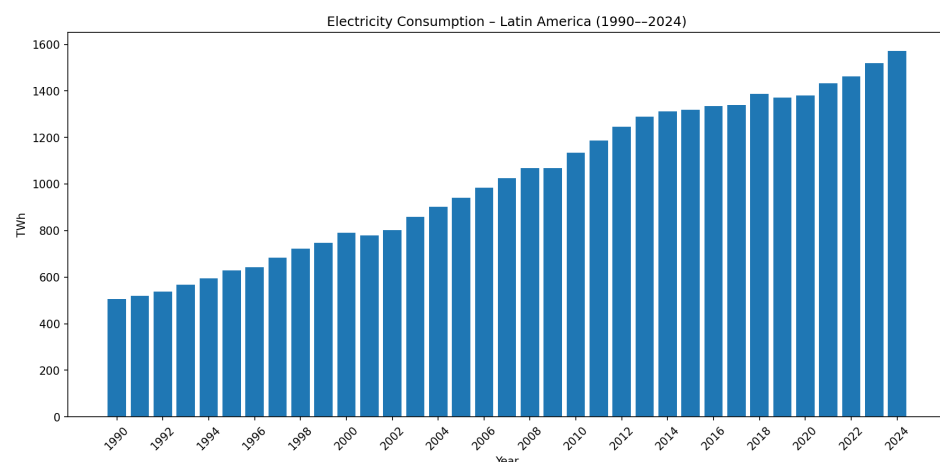


Figure 2. Electricity consumption in Latin America between 1990 and 2024.

Electricity demand is a variable that is not only dynamic, but also strongly influenced by exogenous factors such as weather conditions, economic activity, social behavior patterns, demographic growth, and exceptional events (such as heat/cold waves or pandemics), among others [1,6,7]. Inefficient management of this demand can compromise system security and electricity supply, generate significant economic losses, and even affect essential services such as hospitals, transportation infrastructure, drinking water systems, and data centers [6,8,9].

An inaccurate forecast can have critical consequences for power system operation. Underestimating demand may require the dispatch of more expensive or less efficient generation resources, thereby increasing electricity generation costs [10,11]. Conversely, overestimating demand can result in excess generation capacity, leading to operational

losses, including additional generation and plant availability costs, as well as potential environmental impacts associated with the inefficient use of energy resources [6,12].

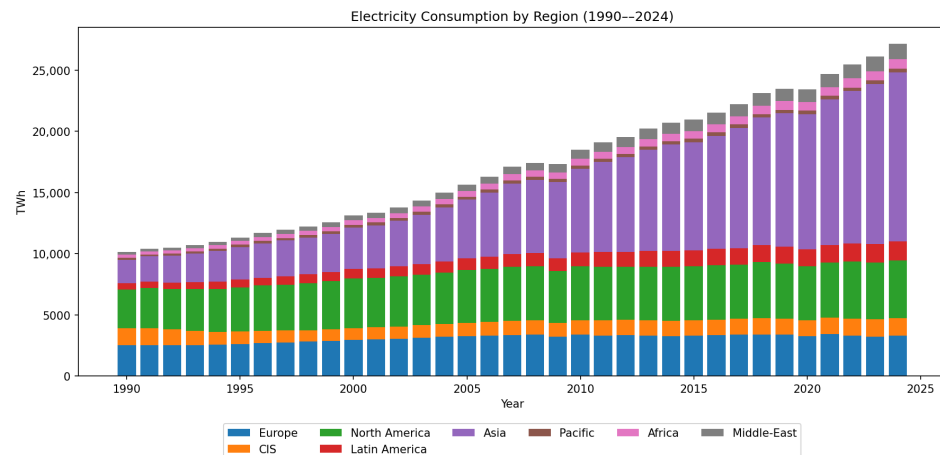


Figure 3. Regional comparison of global electricity consumption between 1990 and 2024.

Furthermore, the increasing integration of renewable energy sources, particularly solar and wind, into distributed generation systems introduces an additional layer of complexity to forecasting due to their intermittent nature. This variability requires greater forecasting accuracy to avoid imbalances between electricity generation and demand, which may compromise grid reliability [6,8,9].

From an operational perspective, demand forecasting is usually divided into three time horizons: short term (from hours to one week), useful for scheduling and operational dispatch; medium term (one week to one year), oriented toward maintenance planning, tariffs design, or contract management; and long term (more than one year), essential for investment decisions in generation and transmission infrastructure [2,10].

Historically, electricity demand forecasting has been addressed using classical statistical approaches, including linear regression and time-series models such as ARMA, ARIMA, and SARIMA. These methods enable the identification of historical consumption patterns and the representation of seasonal behavior and long-term trends [13,14]. However, these models exhibit significant limitations when dealing with data characterized by nonlinear relationships, extreme events, or consumption patterns that deviate from normal operating conditions as a result of technological or climatic changes [3,15].

With the advancement of artificial intelligence (AI), approaches based on *Machine Learning* (ML) and *Deep Learning* (DL) have gained considerable attention in the literature for addressing electricity demand forecasting problems. Their popularity stems from their ability to process large volumes of data, identify complex patterns, model nonlinear relationships, and adapt to highly variable operating conditions [8,16–18]. Algorithms such as *Random Forest*, *Support Vector Machines*, and deep neural networks (DNNs, CNNs, and LSTMs), as well as hybrid models, have demonstrated significant improvements in forecasting accuracy compared with traditional statistical methods. Reported error reductions range from approximately 2% to 80%, depending on the forecasting problem, evaluation metric, and application context. Nevertheless, these methods still face challenges related to overfitting, interpretability, and computational complexity [2,8].

In particular, hybrid models that combine statistical techniques with ML or DL algorithms have been proposed as an intermediate solution that makes it possible to leverage the explanatory capability of traditional models and the predictive capability of *Machine Learning* models [2,16,18]. These approaches integrate the best of both worlds, although they can also increase implementation complexity and the risk of overfitting if they are not properly designed.

In this context, this article provides a comprehensive analysis of the main approaches used for electricity demand forecasting, evaluating statistical models, *Machine Learning*, *Deep Learning*, and hybrid architectures. Their methodological foundations, predictive capabilities, advantages, and limitations are assessed, and the role of exogenous variables in improving predictions is reviewed. In addition, a comparison among models is presented based on criteria such as accuracy, adaptability, scalability, and computational performance.

Unlike existing review studies, which mainly focus on comparing forecasting models through reported error metrics or analyze specific model families independently [2,19], this work proposes a structural taxonomy for electricity demand forecasting models based on three complementary dimensions: methodological family, forecasting horizon, and required data type. This classification framework enables a systematic identification of the relationships, differences, and complementarities among forecasting approaches, while also providing a rigorous basis for comparative evaluation under multiple criteria. In addition, this review explicitly examines emerging architectures such as *Mixture of Experts* (MoE), probabilistic forecasting methods, and advanced hybrid frameworks. These approaches are analyzed within the broader context of the operational challenges introduced by the energy transition, including smart grid expansion, concept drift, increasing uncertainty in electricity demand, and the growing penetration of distributed energy resources.

The remainder of this paper is organized as follows. Section 2 introduces the formal definition of the electricity demand forecasting problem, including its mathematical formulation and the classification of forecasting horizons. Section 3 describes the methodology and evaluation framework adopted for the comparative analysis. Section 4 presents the bibliometric analysis of the reviewed literature. The different forecasting model families are then examined in detail. Section 5 reviews statistical approaches, Section 6 discusses *Machine Learning* models, Section 7 analyzes *Deep Learning* architectures, and Section 8 presents hybrid forecasting approaches. Subsequently, Section 9 provides the qualitative and quantitative comparative analysis, including the proposed taxonomy and the identification of current research gaps. Finally, Section 10 summarizes the main conclusions of this work and outlines potential directions for future research.

2. Electricity Demand Forecasting: Problem Formulation

2.1. Formal Definition

Let $\{y_t\}_{t=1}^T$ denote a univariate time series representing electricity demand (in kWh) observed at hourly intervals. The forecasting problem consists of estimating the future values $\hat{y}_{T+1}, \hat{y}_{T+2}, \dots, \hat{y}_{T+H}$ using all information available up to time T . Formally, the problem can be expressed as:

$$\hat{\mathbf{y}}_{T+1:T+H} = f(\mathbf{y}_{T-L+1:T}, \mathbf{X}_{T-L+1:T+H}) + \varepsilon \quad (1)$$

where H represents the forecasting horizon, L is the length of the observation window, $\mathbf{X} \in \mathbb{R}^{n \times p}$ denotes the matrix of exogenous variables, and ε corresponds to the error term [2,4].

A fundamental distinction in operational practice is between *point forecasting*, which produces a single estimate $\hat{y}_{T+h}$, and *probabilistic forecasting*, which characterizes the conditional distribution $P(y_{T+h} | \mathcal{I}_T)$, where $\mathcal{I}_T$ denotes the full information set available at time T . Probabilistic forecasting is increasingly required in risk-sensitive contexts such as reserve scheduling, energy-market clearing, and demand response activation, as it provides explicit uncertainty quantification for cost-aware decision-making [19,20].

The most commonly used metrics for evaluating point forecasting models are:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (2)$$

$$\text{MAPE} = \frac{100}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (4)$$

where y_i and $\hat{y}_i$ correspond to the observed and predicted values at evaluation step i , respectively, and N denotes the total number of evaluated observations. For probabilistic forecasting, calibration and sharpness are commonly assessed using the Continuous Ranked Probability Score (CRPS) and the Winkler Index [20].

2.2. Forecasting Horizons

Electricity demand forecasting can be categorized into three main temporal horizons, each associated with different operational and planning tasks within power systems [2,21]. Table 1 summarizes their principal characteristics.

Table 1. Classification of electricity demand forecasting horizons.

Horizon	Range	Resolution	Primary Application
STLF	1 h–2 wk	Hourly	Dispatch, markets, reserves
MTLF	1 wk–1 yr	Daily/monthly	Maintenance, contracts
LTLF	1–20+ yr	Monthly/ann.	Expansion, investment, policy

2.2.1. Short-Term Forecasting (STLF)

Short-Term Load Forecasting covers horizons ranging from 1 h to approximately 1–2 weeks, typically with hourly resolution. This forecasting horizon is operationally critical because it directly supports economic dispatch, wholesale electricity market operation, and operating reserve scheduling. In this context, calendar-related variables and meteorological forecasts are among the most influential predictors, particularly in climate-sensitive systems.

2.2.2. Medium-Term Forecasting (MTLF)

Medium-Term Load Forecasting spans forecasting periods from 1 week to 1 year and commonly operates at daily or monthly resolution. Typical applications include maintenance scheduling for generation units, fuel management, and bilateral contract planning. At this temporal scale, economic indicators such as GDP (Gross Domestic Product) and industrial activity, together with seasonal climate patterns, become more relevant than short-term demand dynamics [14,22].

2.2.3. Long-Term Forecasting (LTLF)

Long-Term Load Forecasting addresses horizons ranging from 1 year to more than 20 years ahead. It plays a central role in capacity expansion planning, transmission and distribution infrastructure investment, and energy policy development. At this level, electricity demand is strongly influenced by macroeconomic growth, demographic changes, and the adoption of emerging technologies such as electric vehicles, distributed solar generation, and large-scale data centers [21,23].

2.3. Taxonomy of Solution Approaches

The wide variety of approaches proposed to solve Equation (1) can be broadly grouped into four major methodological families. *Statistical models* are based on parametric assumptions such as linearity, stationarity, and Gaussianity, offering high interpretability and relatively low data requirements. *Machine Learning* models learn flexible nonlinear relationships directly from data without requiring explicit functional assumptions, allowing them to capture more complex demand dynamics.

In turn, *Deep Learning* models employ multiple representation layers to automatically extract hierarchical features from high-dimensional time series data. It is important to note that Deep Learning is a specialized subset of Machine Learning (specifically, the family of algorithms based on multi-layer artificial neural networks) and is treated as a separate category in this taxonomy solely to reflect the distinct research communities, methodological characteristics, and operational trade-offs that differentiate deep architectures from classical ML models in the forecasting literature. Finally, *hybrid models* combine complementary forecasting approaches by assigning different signal components, such as trend, seasonality, and residual patterns, to specialized methods from different model families.

Each methodological family involves different trade-offs in terms of predictive accuracy, interpretability, computational complexity, and data requirements. These characteristics are analyzed in detail in Sections 5–8.

3. Methods and Evaluation Framework

The evaluation and comparison of electricity demand forecasting models requires a structured method that enables their performance to be analyzed under objective and consistent criteria. Given the diversity of existing approaches (from traditional statistical models to advanced methods based on artificial intelligence), it is necessary to establish an analysis framework that considers multiple dimensions: accuracy, robustness, generalization capability, computational efficiency, interpretability, scalability, and suitability for the forecasting horizon.

For this study, a literature review was conducted to identify, select, analyze, and compare representative models used in electricity demand forecasting. The review focuses on publications indexed in academic databases such as IEEE Xplore, and Scopus, prioritizing recent studies (last decade) that report comparable validation metrics.

3.1. Literature Search and Study Selection

This study follows the principal recommendations of the PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework [24] to improve the transparency and reproducibility of the literature selection process. In particular, the search strategy, eligibility criteria, screening procedure, and study-selection workflow are reported explicitly to facilitate independent verification of the review corpus.

3.1.1. Literature Search Strategy

A systematic literature search was conducted using Scopus and IEEE Xplore. Scopus served as the primary database because of its broad multidisciplinary coverage of the major publication venues relevant to electricity demand forecasting, including journals and conference proceedings published by IEEE, Elsevier, MDPI, Springer Nature, ACM, Frontiers, and AIP. IEEE Xplore was used as a complementary source to enhance the coverage of specialized electrical engineering and computer science publications, particularly conference proceedings and IEEE Transactions that may not be fully indexed in Scopus.

To further reduce the risk of omission, backward and forward citation tracking was performed on highly relevant studies identified during the database search, allowing

additional pertinent publications to be incorporated into the review corpus. The search was applied simultaneously to titles, abstracts, and keywords (TITLE-ABS-KEY) using the following Boolean query:

```
("electricity demand forecasting" OR "electricity load forecasting" OR "power demand forecasting" OR "short-term load forecasting" OR "STLF") AND ("machine learning" OR "deep learning" OR "neural network" OR "LSTM" OR "random forest" OR "support vector" OR "transformer" OR "gradient boosting" OR "hybrid model" OR "GRU" OR "mixture of experts")
```

The search was restricted to journal articles and conference proceedings published in English, without imposing additional date or subject-area constraints on the search query. Instead, subject relevance was ensured through the inclusion criteria described in Section 3.1.2. The resulting search strategy was used to identify potentially relevant studies and construct the review corpus. To facilitate reproducibility, the search query, database sources, eligibility criteria, and study-selection workflow are explicitly reported.

3.1.2. Inclusion and Exclusion Criteria

Studies were assessed against the following pre-specified eligibility criteria.

Inclusion criteria:

1. Original primary research that proposes, develops, or evaluates a forecasting model for electricity demand or electrical load prediction.
2. Reports at least one quantitative performance metric (e.g., MAPE, RMSE, MAE, R^2 , nRMSE, or equivalent), evaluated on an independent test set, enabling meaningful cross-study comparisons.
3. Published in English in a peer-reviewed journal or conference proceeding.
4. Conducted within the domains of energy systems, electrical engineering, computer science, applied mathematics, or closely related fields.
5. Can be classified into at least one of the four methodological families considered in this review: statistical models, *Machine Learning*, *Deep Learning*, or hybrid approaches.

Exclusion criteria:

1. Studies focused exclusively on electricity price forecasting or energy market prediction without an explicit electricity demand forecasting component.
2. Studies addressing only supply-side forecasting tasks, such as solar irradiance, photovoltaic generation, or wind power prediction, without modeling electricity demand.
3. Non-primary research publications, including editorials, surveys, review articles that do not propose or evaluate a forecasting model, and dataset description papers without forecasting experiments.
4. Studies that do not report quantitative performance metrics evaluated on an independent test set, thereby preventing reproducible performance assessment and cross-study comparison.
5. Studies for which the full text could not be obtained despite reasonable retrieval efforts.
6. Studies published in a language other than English.
7. Studies in which electricity demand forecasting is not the primary contribution of the work, as determined from the title, abstract, or keywords.

3.1.3. Study Screening and Selection Process

During the screening stage, titles were systematically assessed for relevance to the objectives of this review. Records unrelated to electricity demand forecasting, focused exclusively on electricity price or generation forecasting, or lacking a clearly defined forecasting methodology were excluded. During the eligibility stage, the full text of each potentially relevant study was reviewed to verify methodological completeness,

the availability of quantitative performance metrics, and consistency with the scope of the review.

Eligible studies were subsequently classified into one or more of the four methodological families considered in this work: statistical models, *Machine Learning*, *Deep Learning*, and hybrid approaches. This classification enables a balanced comparative analysis across different modeling paradigms, which constitutes one of the main objectives of the review. Any disagreements regarding the eligibility or classification of borderline studies were resolved through discussion and consensus among the authors.

After applying the inclusion and exclusion criteria and completing the screening and eligibility assessment stages, a final set of 76 studies was retained for comparative analysis. These studies constitute the evidence base used throughout the remainder of this review, and the complete selection process is illustrated in Figure 4.

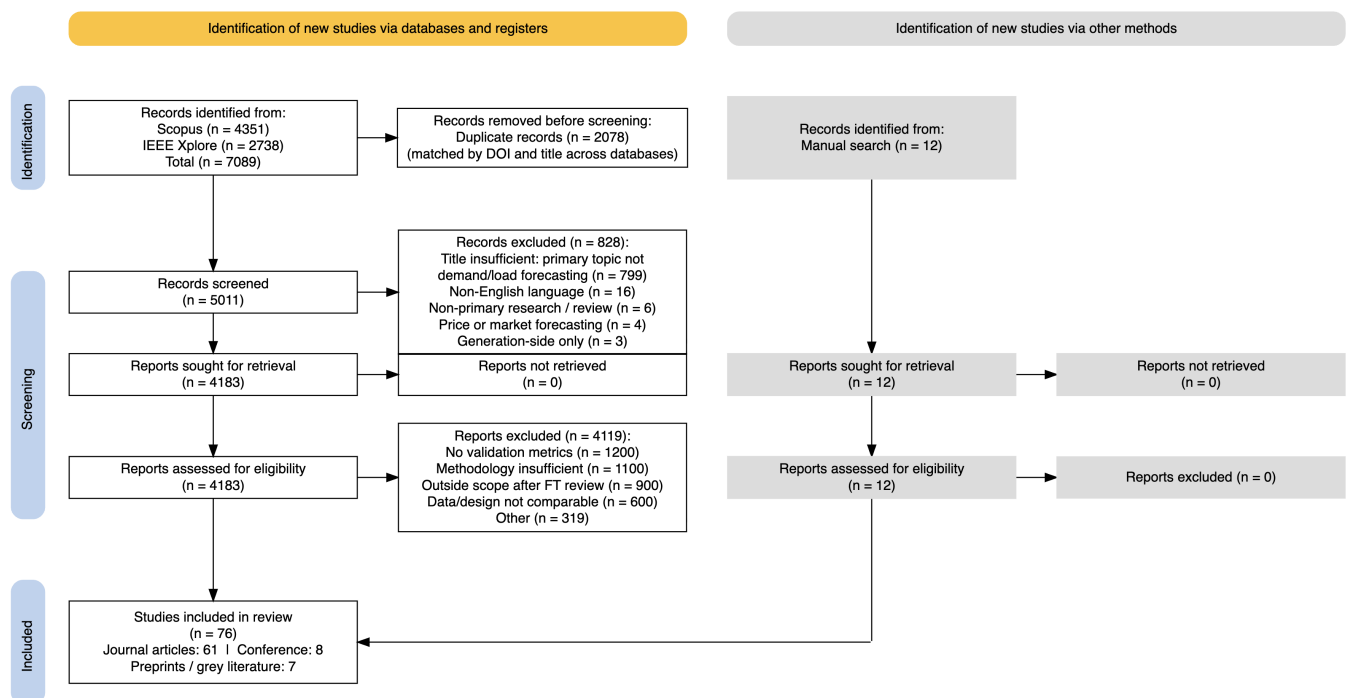


Figure 4. PRISMA 2020 flow diagram of the systematic literature selection process [24].

3.1.4. Data Extraction

For each of the 76 selected studies, the following information was extracted and tabulated: publication details (year, journal or conference, and geographic region); forecasting horizon and temporal resolution; model type and methodological family; exogenous variables used; evaluation metrics and reported accuracy values; and dataset source. Because the selected studies differ substantially in dataset scale, geographic region, temporal resolution, aggregation level, preprocessing pipeline, baseline selection, and validation protocol, a formal quantitative meta-analysis was not methodologically appropriate. Accordingly, this review adopts a narrative synthesis approach to identify performance patterns, strengths, and limitations across model families, as presented in Sections 5–8 and synthesized in Section 9.

3.2. Comparison Criteria

The comparison among forecasting models is structured according to the following key criteria:

- **Forecast horizon:** Models are grouped according to the time horizon for which they were designed: short term, medium term, and long term. This parameter directly affects the type of data required, model complexity, and operational applicability.
- **Performance metrics:** To evaluate model accuracy, standard metrics widely used in the literature are considered, such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). These metrics quantify the difference between predicted and actual values, facilitating comparison among different approaches.
- **Modeling capability and robustness:** The model's ability to capture nonlinear relationships between demand and exogenous variables (temperature, humidity, calendar effects, Gross Domestic Product (GDP), among others) is assessed, as well as its robustness to noise, measurement errors, and outliers in the data, maintaining stable performance under such perturbations.
- **Adaptability and scalability:** The ease with which models can be adjusted to new conditions or different contexts is examined, as well as their behavior when faced with high-dimensional or large-scale datasets.
- **Interpretability:** Although complex models such as deep neural networks offer high precision and accuracy, they often sacrifice interpretability. This parameter analyzes how understandable the results are and how useful the model is as a decision-making tool.
- **Computational cost:** The practical implementation of forecasting models depends not only on their accuracy and precision, but also on their training, updating, and execution costs. This parameter is especially relevant in systems with limited computational resources or that require real-time processing.
- **Handling of exogenous variables:** The type and number of exogenous variables incorporated into the models is considered. More advanced models integrate meteorological, socioeconomic, and behavioral information to improve their predictions.

These parameters not only enable a quantitative comparison between forecasting models, but also provide insight into their suitability for specific applications. In practice, no single model consistently outperforms all others across every evaluation criterion or forecasting scenario. The performance of a forecasting approach depends on multiple factors, including the characteristics of the available data, the forecasting horizon, and the operational requirements of the power system. Consequently, selecting an appropriate model requires balancing aspects such as predictive accuracy, interpretability, adaptability, and computational cost according to the objectives of the application.

Before presenting the analysis of forecasting models, Section 4 introduces a bibliometric characterization of the reviewed literature. This analysis provides an empirical basis for the organizational structure adopted throughout this work.

4. Bibliometric Analysis

To complement the methods and evaluation framework described in the previous section and provide a quantitative characterization of the current state of the field before addressing the qualitative and quantitative analysis of the models, a bibliometric analysis was conducted on the corpus of publications reviewed using VOSviewer version 1.6.20. This analysis makes it possible to identify dominant thematic structures, temporal research trends, and scientific collaboration networks, thereby providing the reader with a panoramic view of the field.

4.1. Thematic Structure and Temporal Evolution

Figure 5 presents the keyword co-occurrence network, where each node represents a term, node size is proportional to frequency of occurrence, and colors distinguish the thematic clusters identified by VOSviewer. Figure 6 shows the temporal evolution of the same network, where the color scale indicates the average publication year: purple tones (2018), blue-green (2020–2022), and green-yellow (2023–2024). The joint analysis of both visualizations makes it possible to characterize both the field structure and its evolutionary dynamics.

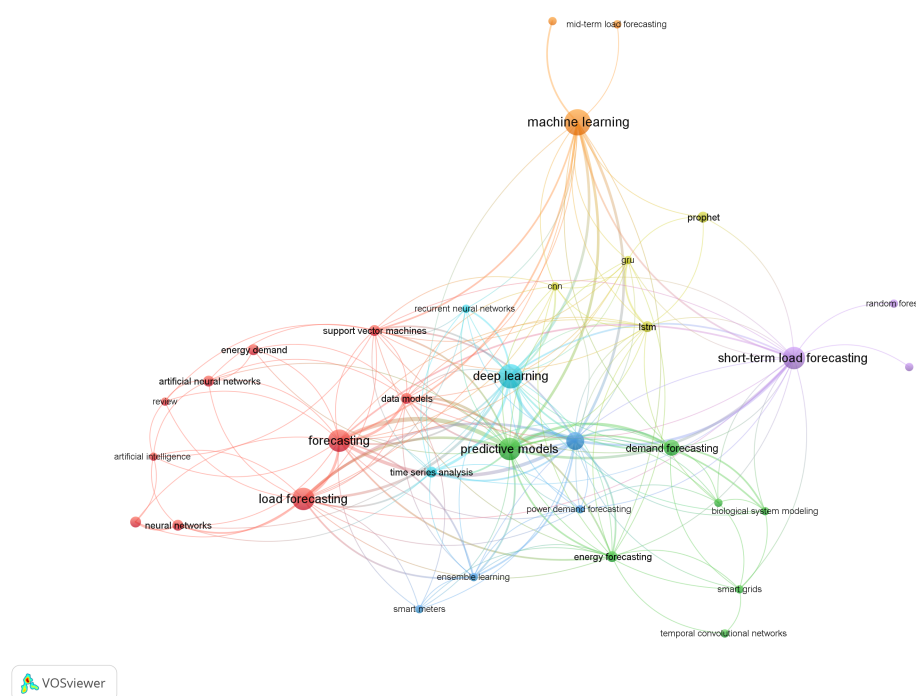


Figure 5. Keyword co-occurrence network generated with VOSviewer. Nodes represent key terms, node size is proportional to frequency of occurrence, and colors distinguish the identified thematic clusters.

The first cluster visible in Figure 5 (red, left side) groups terms such as *forecasting*, *neural networks*, *artificial neural networks*, *support vector machines* (SVM), *energy demand*, *artificial intelligence*, and *review*. This cluster represents the foundation of electricity demand forecasting, where classical neural networks and SVMs constituted the first artificial intelligence approaches applied to the problem. The temporal visualization in Figure 6 confirms this interpretation, since the nodes in this cluster are located in earlier years (2018), indicating that scientific output centered on these methods reached its highest activity before 2020.

The second cluster in Figure 5 (blue, central area) is organized around *deep learning*, *power demand forecasting*, *recurrent neural networks*, *ensemble learning*, *smart meters*, and *time series analysis*. The *deep learning* node is the central node of the entire network, reflecting its role as a unifying concept across the different families. The grouping of *deep learning* with *time series analysis* in this cluster indicates that deep learning has consolidated itself as one of the central axes of time-series analysis in the energy field. In Figure 6, these terms are located in an intermediate time window (2020–2022).

The third cluster in Figure 5 (green, lower-right area) groups *predictive models*, *demand forecasting*, *energy forecasting*, *smart grids*, *temporal convolutional networks*, and *biological system modeling*. This grouping highlights the convergence between modern predictive techniques and smart-grid infrastructure. In the temporal visualization (Figure 6), these

terms are concentrated in the most recent years (2023–2024), indicating an active shift toward these applications.

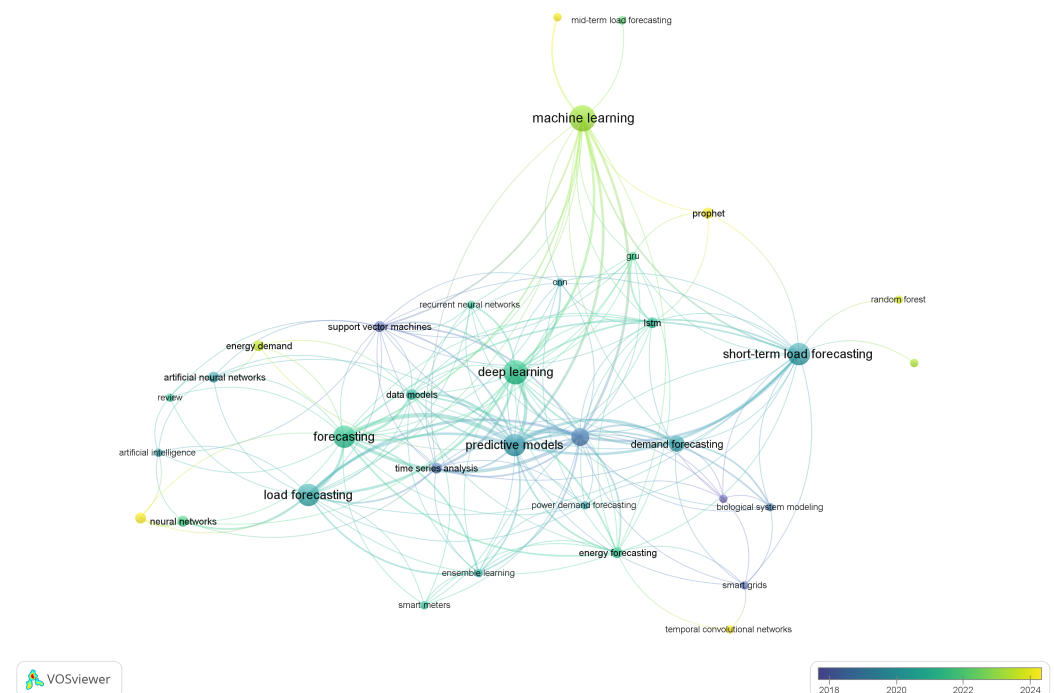


Figure 6. Temporal overlay visualization of the keyword co-occurrence network. The color scale indicates the average publication year: purple tones (2018), blue-green (2020–2022), and green-yellow (2023–2024).

The fourth cluster in Figure 5 (orange, upper area) contains *machine learning* and *mid-term load forecasting*. It is significant that *machine learning*, despite being one of the largest nodes in the entire network, forms a cluster separate from *deep learning*. This separation reflects the operational distinction between both methods in electricity demand forecasting literature: while *deep learning* is associated with time-series analysis and predictive models (blue cluster), *machine learning* appears linked to medium-term horizons and to a more general use of the term. The yellow tone of both nodes in the temporal visualization (Figure 6) indicates that they correspond to the most recent publications (2023–2024). The presence of *mid-term load forecasting* in this cluster, with a similarly recent tone, indicates growing interest in a horizon historically less explored than short-term forecasting, constituting an area with development opportunities.

The fifth cluster in Figure 5 (purple, right area) groups *short-term load forecasting* and *random forest*. The *short-term load forecasting* node is one of the largest in the network, reflecting the concentration of research on the short-term horizon. Its direct association with *random forest* in an independent cluster suggests that this ensemble model has positioned itself as a reference tool particularly for this time horizon, consistent with the evidence analyzed in Section 6.

The sixth cluster in Figure 5 (turquoise, center-right area) concentrates deep learning architectures specialized in sequences: *LSTM*, *GRU*, *CNN*, and *prophet*. The separation of these specific architectures from the generic *deep learning* node (blue cluster) reveals a level of thematic specialization, since while *deep learning* functions as a general term, publications employing *LSTM*, *GRU*, or *CNN* tend to co-occur with one another, forming a differentiated technical community. The inclusion of *prophet* in this cluster indicates that this model, although statistical in nature, has frequently been used in combination with deep architectures in hybrid approaches, as analyzed in Section 8. Temporally, these nodes

are located in the intermediate period (2020–2022), corresponding to their consolidation phase as standard tools for short-term forecasting.

4.2. Co-Authorship Networks

Figure 7 presents the co-authorship network, where each node represents an author and links indicate collaborations in joint publications. Colors distinguish the collaboration clusters identified by VOSviewer, and the size of each node is proportional to the author's number of publications in the analyzed corpus.

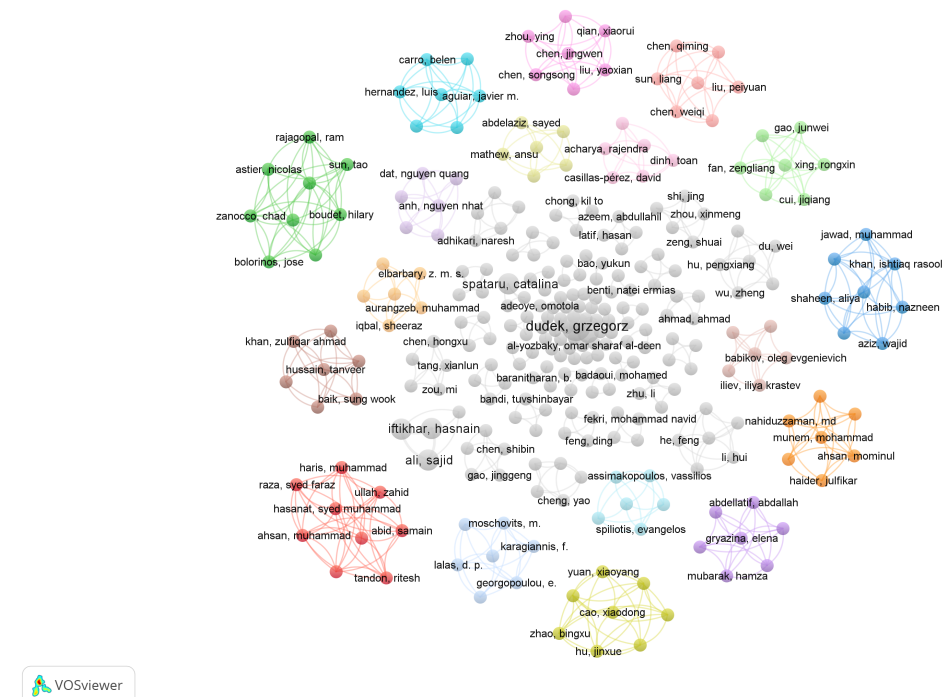


Figure 7. Co-authorship network generated with VOSviewer. Each node represents an author, links indicate joint publications, and colors distinguish the identified collaboration clusters.

Figure 7 reveals a fragmented collaboration structure, composed of multiple small and medium-sized clusters with few inter-cluster links, where groups work relatively independently and large transnational collaboration networks have not yet consolidated. The author with the highest centrality and publication volume is Dudek, G., whose contributions range from *Random Forest* models to probabilistic *stacking* techniques for short-term forecasting [20,25,26].

Several geographically distributed collaboration clusters are identified in Figure 7. In the upper area, a group of European and Latin American authors is observed, including Hernández, L., Aguiar, J.M., and Carro, B., associated with research on forecasting in smart grids and *microgrids* [12]. In the central and left area, a set of authors from Asia and the Middle East forms clusters that are densely connected internally, with researchers such as Iftikhar, H., Ali, S., and Khan, Z.A., who have contributed predominantly to studies using LSTM, SVR, and *boosting* techniques [13,14,27]. In the lower area, another group includes Greek and European authors such as Spiliotis, E., Assimakopoulos, V., and Karagiannis, F., linked to work on advanced statistical models and probabilistic forecasting [28,29].

The fragmentation observed in the co-authorship network has direct implications for progress in the field. The lack of interaction among groups working on statistical, *Machine Learning*, and *Deep Learning* approaches may limit the development of truly integrative hybrid models, which require multidisciplinary expertise. This finding reinforces the need to foster interdisciplinary collaborations among currently disconnected research

communities, particularly between groups specialized in statistical techniques (with strong interpretability) and those focused on deep architectures (with high predictive capacity), a topic revisited in Section 9 and in the future research lines of this article.

Taken together, the thematic cluster structure and co-authorship patterns identified in this bibliometric analysis provide empirical support for the organizational structure adopted in the following sections. Cluster 1 (red), which is primarily associated with classical neural networks and support vector machines applied to load forecasting, establishes the basis for the review of *Machine Learning* models presented in Section 6. Cluster 2 (blue), focused on *deep learning* and time-series analysis, motivates the discussion developed in Section 7. In addition, Cluster 6 (turquoise), which groups architectures such as LSTM, GRU, CNN, and Prophet due to their frequent co-occurrence in the literature, highlights the growing relevance of hybrid forecasting approaches examined in Section 8.

Furthermore, the fragmentation observed in the co-authorship network suggests that research in electricity demand forecasting remains distributed across partially disconnected communities. This observation reinforces the importance of integrative and interdisciplinary approaches, which are further discussed in the comparative analysis presented in Section 9.

5. Statistical Methods

As suggested by the bibliometric analysis (Section 4), statistical approaches constitute the foundational cluster of the field, with the longest research trajectory and highest interpretability.

Traditional models for electricity demand forecasting are based on statistical and mathematical principles to analyze historical electricity demand patterns and predict future demand. These models seek to identify trends, seasonality, stationarity, and relationships with explanatory variables in order to generate future demand predictions. Several of these approaches used for electricity demand forecasting are presented below.

Statistical models have historically been the cornerstone of electricity demand forecasting, particularly at a time when data availability was limited and electrical systems were relatively simpler in their operational structure. These models, grounded in mathematical assumptions and linear relationships, make it possible to establish explicit links between observable variables and expected outcomes, which has facilitated both their application in practical settings and their understanding by system operators. In a context where transparency and interpretability are essential, statistical models have provided a solid foundation for the development and evolution of more complex approaches.

One of the most widely used statistical approaches has been linear regression, which has been extensively applied in multivariate models that seek to establish relationships between electricity demand and socioeconomic or climatic variables. For example, in ref. [1], multivariate regression was used to forecast short-term electricity demand using Gross Domestic Product (GDP) and population growth as explanatory variables, together with historical consumption data. This model achieved a MAPE of 4.13% with respect to actual values, reflecting good accuracy for short-term horizons. However, when the same model was applied to medium-term scenarios, the error increased to 55%, revealing a structural weakness of this approach: its limited capacity to adapt to dynamic patterns, nonlinear behaviors, or unexpected external factors. This weakness has been widely documented in the literature and constitutes a central limitation of models based on strictly linear relationships [9].

Improvements in the accuracy of regression models have been possible through the incorporation of explanatory or exogenous variables. In ref. [30], for example, a multivariable regression model is proposed for monthly electricity demand forecasting in South Korea, covering a dataset from 2012 to 2023. This study incorporated variables such as climate,

calendar effects, industrial activity, and unexpected events such as the COVID-19 pandemic. The results showed that, even against more sophisticated models such as Holt–Winters (a type of exponential smoothing model), and even against machine learning and deep learning models, the regression-based approach remained highly competitive. This finding is especially important because it shows that the key is not necessarily model complexity, but rather the quality, relevance, and treatment of the independent variables used. As an explanatory method, regression also makes it possible to directly interpret the influence of each variable on demand, which provides a significant advantage in terms of transparency and interpretability.

In other studies such as [23], the linear regression approach was used to project non-residential energy demand in 14 West African countries between 2000 and 2030. Despite the simplicity of the method, the authors reproduced demand with errors below 15% and projected growth from 55 TWh to 286 TWh, validating its use for energy planning in low-data-availability contexts. This type of study shows that linear regression can be useful for establishing “top-down” models, which are characterized by analyzing overall energy system behavior from aggregated macroeconomic variables, such as GDP, population, or level of industrialization, to infer the evolution of total energy demand. In this approach, reasoning proceeds ‘from top to bottom’, that is, from the economic and structural determinants of the system toward its components or sectors, allowing the relationship between economic growth and energy consumption to be captured without requiring detailed technological or sector-level information.

Regression models have also been extended with autoregressive structures, seasonal components, or auxiliary binary variables (known as dummy variables) to incorporate holidays and extreme climatic events. For example, in ref. [28], a daily and monthly model was developed for Greece between 1993 and 2002, integrating temperature, humidity, and heating degree days (HDD) and cooling degree days (CDD), together with seasonal components. By doing so, a maximum error of only 2.7% was achieved in the monthly model. This approach shows that even with statistically conventional techniques, high levels of accuracy can be achieved if the model is well specified and supported by adequate data.

Similarly, in ref. [21], an extended regression model was applied to the state of Texas, a region characterized by rapid growth in AI-driven data centers. By incorporating not only climatic variables, but also annual lags and categorical variables to represent reordered months, the model reached a MAPE of 2.10%, outperforming even advanced models such as XGBoost (Extreme Gradient Boosting), LightGBM (Light Gradient Boosting Machine), and LSTM (Long Short-Term Memory). This result reinforces the idea that, in certain contexts, a well-designed regression model can outperform more complex models, especially when structural relationships in the data are well understood and represented.

Beyond regression, time-series models have occupied a central place in energy demand forecasting, since they allow temporal properties of data, such as trend, seasonality, and autocorrelation, to be exploited. Among the most common are ARIMA (Autoregressive Moving Average), SARIMA (Seasonal ARIMA), and SARIMAX (Seasonal ARIMA with exogenous variables), all widely documented in the literature [2,12,31]. These models are particularly effective for stable and homogeneous series, as shown in ref. [1], where an additive time-series model was applied to forecast residential demand in Colombia, achieving an error of only 0.87%. Likewise, Refs. [13,14] show the effectiveness of these models in environments such as Pakistan and the Nord Pool market, reporting deviations on the order of 1.9% when decomposition techniques are used to separate deterministic and stochastic components.

One of the most relevant advances has been the incorporation of exogenous variables and cross effects, which has allowed SARIMAX models to be adapted to more complex contexts. In ref. [32], for example, an enhanced SARIMAX model was used for forecasting in Japan between 2012 and 2015, incorporating climatic variables such as temperature and humidity, and calendar factors such as hour, day, and holidays. The model outperformed its standard version by 22% in terms of accuracy, and even improved performance relative to linear regression with independent variables and a basic ARIMA model. This approach is relevant because it shows that even within classical models, levels of sophistication can be introduced that allow them to compete with more modern and advanced techniques.

In ref. [33], an online extension of the SARIMA model, called SARIMA-ONS, was proposed for real-time data processing. Applied in Vietnam, this methodology demonstrated training times three times faster than classical SARIMA and achieved an average MAPE of 4.57%. This version represents a natural evolution of the statistical approach by adapting it to the dynamic needs of smart electrical systems.

Despite their many strengths, statistical models have significant limitations when faced with data exhibiting high nonlinearity, multiple sources of variability, or high penetration of intermittent renewables. In these contexts, their adaptability and predictive power tend to be lower than AI-based models [34]. Even so, several studies show that they can play a complementary or even competitive role, especially when interpretability, low computational cost, or the need to make regulatory decisions based on transparent and auditable models is prioritized [21,28,35].

In addition, probabilistic hierarchical approaches such as the one proposed in [36] have shown improvements in accuracy and hierarchical coherence using smart-meter data in the United Kingdom, introducing coherent probabilistic forecasts at multiple aggregation levels. Likewise, probabilistic approaches based on additive quantile regression have been applied to capture uncertainty in smart-meter data.

Finally, in ref. [37], a weekly peak-demand forecasting framework was developed for Mexico (2017–2024), combining spectral analysis and time-series models. In that study, a SARIMA model reinforced with spectral analysis achieved a MAPE of 1.73%, outperforming models such as Prophet (2.88%) and XGBoost (2.19%) in several regions, highlighting the robustness of enhanced classical approaches against more complex alternatives.

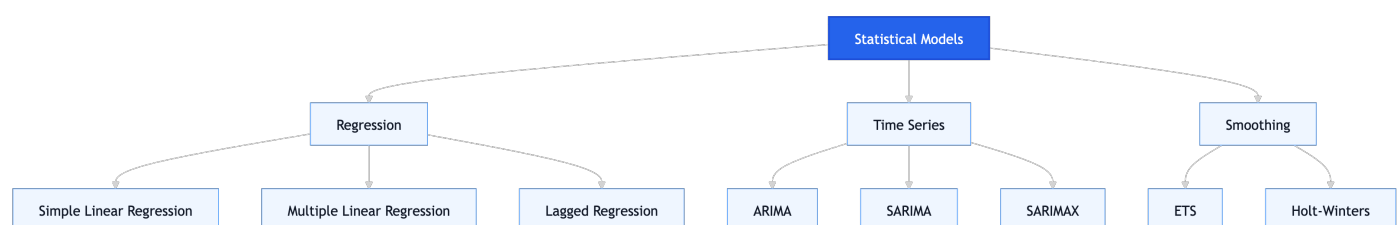
In summary, statistical models, particularly those based on multivariate regression and time series with exogenous variables, remain valid and effective tools for electricity demand forecasting. Their performance depends largely on data quality, model suitability to context, and integration of relevant variables. Although their prominence has been partially displaced by machine learning models, their value lies in interpretability, simplicity, and capacity to generate robust predictions when applied with technical rigor and system knowledge.

The limitations identified in this section, particularly the difficulty of capturing non-linear relationships and adapting to highly variable environments, motivate the transition toward *Machine Learning* models discussed in Section 6. Unlike traditional statistical approaches, these models relax strict linearity assumptions and can exploit more complex feature representations to model electricity demand dynamics more effectively.

The models reviewed in this section can be grouped according to their main characteristics, advantages, and limitations, as summarized in Table 2, which provides a detailed comparative summary of the statistical models analyzed. Complementarily, Figure 8 organizes this family into a visual taxonomy that highlights the relationship between regression and time-series models.

Table 2. Description of statistical models for electricity demand forecasting.

Model	Typical Application	Key Advantages	Limitations
Simple linear regression	Modeling direct relationships between a single explanatory variable (such as temperature) and demand.	Very easy to interpret; useful for understanding univariate sensitivities. Ideal as a baseline or validation control.	Does not capture multiple effects or nonlinear relationships. Models only one causal direction.
Multiple linear regression (MLR)	Estimating demand from multiple exogenous factors (weather, calendar, prices).	Good explainability; allows quantifying the impact of each independent variable. Scalable to multiple predictors.	Assumes strictly linear relationships; sensitive to multicollinearity; less flexible under complex dynamics.
Lagged multiparametric regression	Hourly or intraday forecasting using current and lagged meteorological variables.	Increases accuracy with simple structures; useful in environments with low computational capacity.	Does not capture complex nonlinearities; requires careful feature engineering.
ARIMA (AutoRegressive Integrated Moving Average)	Short-term forecasting on stationary univariate series, without requiring external variables.	Captures the internal temporal dynamics of the series; adapts to smooth trends; performs well with limited data.	Does not handle external variables or multiple seasonalities; requires prior transformation for non-stationary data.
SARIMA (Seasonal ARIMA)	Daily or monthly electricity demand forecasting with repetitive seasonal patterns.	Incorporates explicit seasonality through additional components; useful for data with regular cycles.	Complex to parameterize; poor performance if patterns change dynamically. Does not integrate external effects well.
SARIMAX (SARIMA with exogenous variables)	Demand modeling with multiple seasonalities and external factors (weather, calendar).	Integrates exogenous variables and cross effects; good accuracy and transparency.	Requires high domain knowledge for variable design. Sensitive to errors in external predictors.
Exponential smoothing (ETS, Holt–Winters)	Smoothing and projection of series with trend and seasonality.	Simplicity; useful for data without strong exogenous effects; provides stable and smooth results.	Does not allow direct inclusion of external variables; not suitable for non-stationary systems with abrupt changes.

**Figure 8.** Taxonomy of statistical models for electricity demand forecasting.

6. Machine Learning

Over the last two decades, the emergence and consolidation of Machine Learning (ML) models has represented a turning point in electricity demand forecasting. Unlike classical statistical models, ML approaches do not require rigid assumptions about data distribution or the functional form of relationships between variables, which allows them to adapt to complex, nonlinear, and dynamic environments. This flexibility, combined with the growing availability

of data from smart meters, SCADA systems, and external sources such as meteorological stations, has driven the adoption of ML techniques in power systems worldwide.

ML models are particularly effective when facing prediction problems in which the relationship between independent variables and the target variable is not evident, is highly nonlinear, or is subject to multiple cross interactions. These are precisely the characteristics that define electricity demand behavior, especially in scenarios with high renewable penetration, significant climatic influence, and structural changes in consumption patterns, such as those caused by digitalization, electric mobility, or decentralized generation.

One of the most widely used ML approaches for demand forecasting is Support Vector Machines (SVM), particularly its regression variant known as Support Vector Regression (SVR). These models have shown strong performance in problems where data dimensionality is high and a decision function robust to noise and outliers is required. In ref. [12], a comparison among SVM models, artificial neural networks, and linear models concluded that nonlinear approaches such as SVM offered better accuracy in disaggregated and highly volatile environments. This finding was reinforced in [27], where SVR was used to model demand in a substation in Muzaffarabad, Pakistan, incorporating meteorological variables. The model not only outperformed methods such as *k-Nearest Neighbors* (*k*-NN) and *Random Forest*, but also achieved estimated savings of 0.313 million PKR per day in production costs by reducing MAPE by 2.22%. This ability of SVMs to adapt to real-world problems with high operational variability positions this type of model as a useful tool in contexts where high short-term accuracy is required.

Hyperparameter optimization has been another key aspect in improving model performance. In ref. [38], SVR was applied together with metaheuristic optimization techniques on data from the United States and China, achieving a MAPE of 1.35%, outperforming both neural-network models and ARIMA. Similarly, in ref. [39], a hybrid approach was proposed in which SVR initializes a radial basis function neural network (RBFNN), optimized through an extended Kalman filter, reaching notably low error levels (MAPE of 0.56–1.03%) across diverse operating scenarios. These studies show that SVR effectiveness lies in configuration, training, and fine hyperparameter tuning, which can be decisive for achieving competitive performance against more complex *Machine Learning* or *Deep Learning* models.

On the other hand, decision-tree-based models have been widely adopted due to their ability to handle large volumes of heterogeneous data and their ease of interpretation. The *Random Forest* (RF) model, a bagging-type ensemble model that combines multiple trees trained in parallel over random subsets of the dataset, has become one of the most robust and reliable tools in forecasting tasks. In ref. [40], RF and decision trees were applied to model energy demand in Turkey, considering exogenous variables such as GDP, meteorological data, and trade flows. The model integrating all variables obtained a MAPE of 6.5%, standing out for computational simplicity and good generalization capability. However, the study also showed that compared with more advanced ensemble methods, RF could lag behind when interactions among variables were highly complex.

The potential of Random Forest has been widely validated in studies spanning multiple regions and temporal horizons. In ref. [41], an RF model optimized with Bayesian search was proposed for ultra-short-term forecasting (15 min) in China, reaching a MAPE of 0.161% and an R^2 of 0.979, outperforming models such as XGBoost and LightGBM. Likewise, Ref. [25] conducted a study in four European countries (Poland, United Kingdom, France, and Germany), evaluating seven pattern representations and three training modes, where RF outperformed both classical statistical models and deep neural networks. This result is especially relevant because it shows that, under certain configurations, tree-based models can compete directly with architectures that are much more computationally expensive, such as LSTM or CNN, while maintaining a high level of interpretability compared with DL approaches.

Another highly relevant variant within machine learning is *Gradient Boosting*, particularly implementations such as XGBoost and AdaBoost (Adaptive Boosting), widely recognized for efficiency and accuracy. Gradient Boosting and XGBoost are *boosting*-type ensemble models that build many small decision trees sequentially, where each new tree focuses on correcting errors from the previous set using the gradient of the loss function. In ref. [42], Gradient Boosting and XGBoost models were compared with hybrid Deep Learning approaches for residential demand forecasting, finding that although *boosting* models obtained solid results (MAE of 0.66, RMSE of 0.35 for the Gradient Boosting model), they were outperformed by hybrid approaches. However, in ref. [43], a study applied to the Russian electricity market showed that XGBoost not only achieved a MAPE of 1.41%, but also improved economic efficiency of dispatch decisions when integrating an economic indicator (BMI), showing that model evaluation should not be limited only to statistical metrics, but should also incorporate financial profitability criteria.

For AdaBoost, Ref. [44] developed a model for medium-term monthly electricity demand forecasting in three Chinese provinces using data from 2007 to 2020. The model showed error reduction of up to 82% with respect to individual models, achieving a MAPE of 0.72%. In addition, the model provides narrow and reliable intervals, positioning it as a reference for forecasting under uncertainty by notably improving accuracy and reliability, although at the cost of lower interpretability.

Furthermore, in contexts with data constraints or high uncertainty, tree-based models have shown high robustness. In ref. [26], *Random Forest* was applied to remove non-stationarity and improve prediction in Poland, achieving a MAPE of 1.16%. In ref. [45], the performance of RF, SVR, and MLP was evaluated for long-term forecasting in England and Wales, finding that tree-based methods offered the best accuracy and scalability over extended horizons. This empirical evidence reinforces the idea that adaptability and resilience make tree-based models a reliable option, especially when combined with robust preprocessing or feature-engineering techniques.

The most recent advances in Machine Learning have evolved toward complex ensemble models, including techniques such as *stacking* and *Mixture of Experts* [46–48]. Stacking-type ensemble models train several base models and then a meta-model learns to combine their predictions. These models combine multiple algorithms to capture different aspects of the data, improving both prediction accuracy and stability. For example, in ref. [20], a probabilistic *stacking* method using *Quantile Regression Forests* was proposed, significantly improving interval calibration and statistical coverage. This type of approach is particularly useful in power systems that require not only point values, but also reliable uncertainty estimates.

Mixture of Experts (MoE) models allow decomposition of the problem into subcomponents handled by specialized “experts.” This approach combines multiple models with a routing model, also called a gating model (an architectural component that dynamically assigns weights to expert models), which determines which expert to activate according to input data conditions or characteristics, increasing precision, flexibility, and interpretability [46]. Unlike traditional ensemble models, MoE enables dynamic and conditional expert assignment, optimizing learning in contexts with different behaviors.

In ref. [47], a Mixture-of-Experts framework with *Federated Learning* was proposed for short-term forecasting in smart electrical grids. Using Ausgrid (Australia) data with 30-min resolution between 2010 and 2013, the model achieved improvements of 6 to 9% on static datasets, 4 to 8% on real-time data streams, and up to 41% in *zero-shot forecasting* scenarios (the ability of a model to generalize predictions with lower error for a time series on unseen data during training), outperforming models such as LSTM, CNN, ResNet, and Transformers, and additionally reducing training time by 18%.

Another use case appears in [46], where an explainable Mixture-of-Experts (MoE) approach was presented for extreme scenarios such as heat waves, cold waves, and holidays. The model decomposes total demand into interpretable components (regular, climate-related, and holiday-related), assigning specialized experts such as XGBoost or Informer to each component. Applied to hourly data from China (2017–2019), it achieved a MAPE of 2.9% on normal days, between 3.1–3.9% under extreme weather conditions, and as low as 1.79% on holidays. This design provides not only higher accuracy compared with Deep Learning models such as LSTM (Long Short-Term Memory) or TCN (Temporal Convolutional Networks), but also greater transparency by detailing the contribution of each source of demand variability.

Similarly, ref. [48] proposed TriForecaster, an architecture based on Mixture of Experts within a Multi-Task Learning scheme, capable of simultaneously addressing regional, contextual, and temporal variations. This approach integrates specialized modules such as RegionMixer (a model that combines information across regions) and CTSpecializer (a context-and-time specialized model built from ContextMoE + TimeMoE), together with a stochastic fusion mechanism and a loss function. Evaluated on four datasets with different granularities, the model achieved a 22% reduction in MSE and a 12% reduction in MAE, outperforming architectures such as LSTM, MLP, and GCN-LSTM. Its ability to scale and adapt to the complexity of regional electrical systems positions it as a state-of-the-art reference for electric load forecasting in multi-region environments.

In summary, Machine Learning models have significantly expanded predictive capacity in the energy sector, enabling the solution of problems where statistical approaches are insufficient. Their main strength lies in the ability to capture complex and nonlinear patterns, adapt to multiple data types, and scale in environments with high information volumes. However, effective implementation requires appropriate variable selection, cross-validation techniques, hyperparameter tuning, and, in many cases, energy-domain knowledge to ensure that models are not only accurate, but also operationally useful.

While *Machine Learning* models considerably improve predictive capabilities beyond traditional statistical approaches, accurately capturing long-range temporal dependencies and learning hierarchical representations from high-resolution data often requires more advanced architectures. These challenges have motivated the adoption of *Deep Learning* frameworks, which are reviewed in Section 7.

The models reviewed in this section can be grouped according to their main characteristics, advantages, and limitations, as summarized in Table 3, which presents a detailed comparative summary of the analyzed Machine Learning models. In addition, Figure 9 presents a taxonomy of Machine Learning approaches, distinguishing base models and ensemble strategies to highlight their differences in complexity.

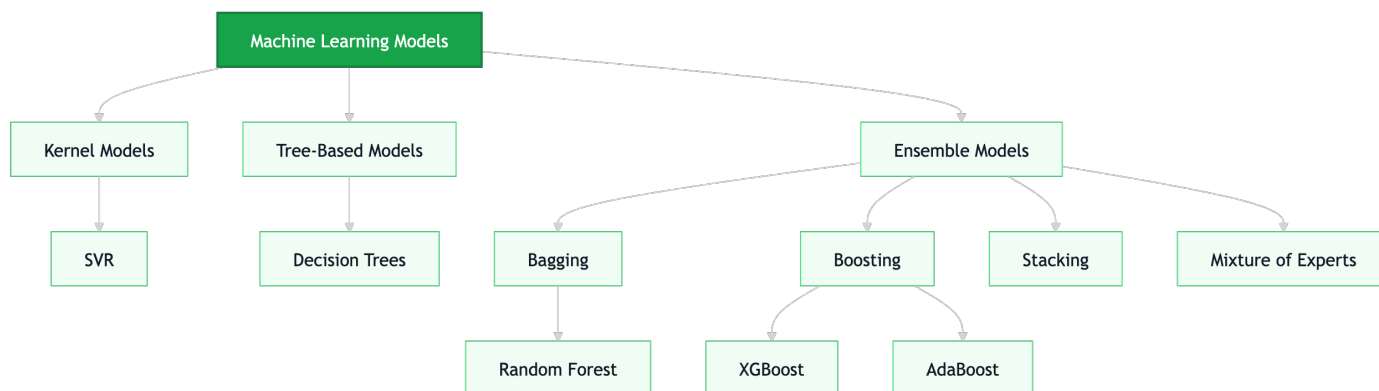


Figure 9. Taxonomy of Machine Learning models for electricity demand forecasting.

Table 3. Description of Machine Learning models for electricity demand forecasting.

Model	Typical Application	Key Advantages	Limitations
<i>Support Vector Regression (SVR)</i>	Short- and medium-term demand prediction where relationships among variables are nonlinear.	Excellent accuracy even with limited data; robust to noise; performs well with Gaussian or polynomial kernels.	Sensitive to hyperparameter selection; difficult to interpret internally; limited scalability.
Decision Trees	Consumption segmentation by customer type, behavior, or location.	High interpretability; useful for visualizing decision rules and as a basis for ensemble models.	Low accuracy as a standalone model; prone to overfitting; unstable under small data changes.
<i>Bagging (Random Forest)</i>	Daily or hourly demand prediction when multiple exogenous variables of different types are available.	Highly robust; handles categorical and numerical data well; avoids overfitting by averaging multiple trees.	Loss of individual interpretability; may require a large number of trees for stability.
<i>Boosting (XGBoost; AdaBoost)</i>	Complex systems where the best possible accuracy is required with multiple, even correlated, variables.	Very high accuracy; efficient handling of missing and categorical data; aggressive computational optimization.	More difficult to tune; higher memory consumption; limited explainability without additional techniques (SHAP, LIME).
<i>Stacking</i>	Combination of several models for final prediction or interval construction.	Improves accuracy and uncertainty calibration; robust under variability.	High complexity; difficult interpretation; requires a diverse set of models.
<i>Mixture of Experts (MoE)</i>	Forecasting under multiple conditions or regions; decomposition by sources of variability.	Adaptive and modular; allows specialized assignment and improves accuracy in heterogeneous or extreme scenarios.	Higher computational complexity; requires expert gating design and appropriate expert specialization; susceptible to expert collapse during training; sensitive to data latency in real-time routing.

7. Deep Learning

Deep Learning (DL) is a subcategory of *Machine Learning* based on artificial neural networks with multiple layers (deep architectures) that enable automatic learning of hierarchical and abstract representations directly from data. Unlike classical *Machine Learning* models, which depend heavily on manual feature engineering, DL models integrate feature extraction and predictive modeling into the same process, allowing them to capture highly complex nonlinear relationships and temporal dynamics that are difficult to represent with traditional approaches. In the context of electricity demand forecasting, this capability is especially relevant for modeling multivariate time series, high-resolution data, and information streams from sensors such as smart meters, where interactions among variables exhibit complex structures and long-range dependencies.

One of the most traditional approaches within DL is the use of artificial neural networks (ANN), which have been used in electricity demand forecasting tasks at both short- and medium-term horizons. In ref. [6], ANN models were compared with multiple-regression and fuzzy-logic models in Turkey using normalized hourly data and lagged variables. Results showed that ANN offered superior accuracy, maintaining lower MAPE and greater consistency, even at horizons up to 168 h. Likewise, in ref. [49], neural net-

works applied to Kerala and Tamil Nadu states in India showed that ANN models with climate variables substantially improved performance in regions with high sensitivity to temperature and humidity, although in regions with dominant industrial consumption the benefits of meteorological variables were less evident.

However, the main leap in energy-forecasting precision and accuracy through DL has come from the introduction of recurrent neural networks (RNN), and especially their advanced variants such as *Long Short-Term Memory* (LSTM) and *Gated Recurrent Units* (GRU). These architectures are specifically designed to process temporal sequences, as they include internal memory mechanisms that allow them to retain relevant information from past events and capture long-term dependencies. This attribute is particularly useful in electricity demand forecasting, where consumption patterns are influenced by daily, weekly, and seasonal behaviors, as well as exogenous events.

In ref. [16], an LSTM-based model enhanced with MIDAS (*Mixed Data Sampling*) was presented, allowing inputs to be weighted according to their volatility. This technique increased model accuracy by better capturing demand fluctuations, showing an advantage over traditional LSTM models. In ref. [10], an LSTM model was implemented for hourly forecasting in Antioquia (Colombia), integrating exogenous variables such as temperature and day type. In that study, the LSTM model outperformed the reference system of the national electric operator (XM), with an average error of 493 MWh in supply over one operating week, demonstrating direct applicability in real contexts.

Other studies have reinforced the superiority of these recurrent architectures over traditional and ML models. In ref. [50], ANN, RNN, and LSTM were evaluated using data from Pakistan and Australia, concluding that LSTM provided the best daily accuracy (MAPE of 1.01%) and weekly accuracy (1.09%), clearly outperforming conventional neural networks. This result was confirmed in [51], where an LSTM-RNN applied to hourly data from Germany (2006–2017) achieved a MAPE of 1.5% for hourly forecasting and 1.77% for annual forecasting, outperforming GRU and CNN while integrating climatic and renewable-generation variables. These findings consolidate LSTM as a high-quality model for high-accuracy electricity forecasting, especially in contexts with strong renewable penetration and high seasonal variability.

Additionally, in ref. [52], PowerLSTM was developed as an optimized model for smart-meter data. This architecture incorporated historical, climatic, and calendar variables, and achieved error reduction above 20% compared with models such as SVR and Gradient Boosting, reaching a MAPE of 4.88%. This study highlighted the importance of considering data granularity and integrating multiple exogenous sources to maximize predictive performance. In environments such as smart grids, where data are abundant but highly dynamic, the ability of LSTM to adapt and generalize becomes a strategic advantage.

Another relevant architecture within DL is the convolutional neural network (CNN), originally developed for image processing but successfully adapted to time-series analysis through one-dimensional convolutions. CNNs are especially effective at capturing repetitive local patterns, such as daily consumption cycles. In ref. [17], the performance of CNN, RNN, and a hybrid CNN-RNN model was compared using data from South Korea. Surprisingly, the standalone CNN outperformed the other configurations, obtaining an R^2 coefficient of 0.992 and an MAE of only 0.02513. This result challenges the common assumption that only recurrent networks are suitable for time series and suggests that, with appropriate architecture and preprocessing, CNNs can deliver equally competitive results.

Building on this foundation, more advanced models have emerged, such as *Temporal Convolutional Networks* (TCN), which use dilation mechanisms to process temporal data sequences [53]. In ref. [54], a TCN applied to household-consumption data in London showed better performance than LSTM in terms of MAE and sMAPE (symmetric mean

absolute percentage error), along with lower computational cost. Likewise, Ref. [55] introduced TCN-QR variants (TCN adapted for quantile regression, i.e., instead of predicting a single value, it estimates quantiles such as P10, P50, or P90 to build prediction intervals and perform probabilistic forecasting) and TCN-MC (TCN with a Monte Carlo method, where random sampling is used to generate multiple predictions and estimate uncertainty and prediction intervals), improving interval coverage and calibration compared with models such as ARIMA or Gaussian processes. Due to their natural parallelization, these architectures are especially useful for real-time applications where computational efficiency is as important as accuracy.

In terms of innovation, hybrid models that combine multiple architectures have shown superior performance compared with individual models. In ref. [42], a CNN-GRU model applied to residential data from the UCI repository [56] (a repository that hosts specialized Machine Learning datasets, owned by the University of California, Irvine) achieved lower MSE and MAE than standalone CNN and LSTM networks. Similarly, Ref. [57] integrated CNN, GRU, and multi-head attention mechanisms (internal model layers that assign weights to different parts of the input sequence to highlight the most relevant information) to improve demand prediction under variable weather conditions, reducing errors by more than 50% compared with traditional models. These hybrid approaches not only leverage CNN local-pattern extraction and GRU sequential memory, but also integrate attention mechanisms that identify the most relevant input segments for each prediction, thus improving model interpretability and adaptability.

Another hybrid-architecture approach was presented in [58], where five configurations were compared for daily electricity demand forecasting in Germany: multiple linear regression, artificial neural networks (ANN), convolutional neural networks (CNN), long short-term memory (LSTM), and a hybrid ANN-LSTM model. The latter, which combines ANN pattern-extraction capacity with LSTM sequential memory, showed the best performance, achieving a MAPE of only 0.89%. This result not only validates the effectiveness of deep networks, but also shows that combining them can generate synergies that outperform each architecture individually. It is worth noting that these configurations, although computationally more demanding, enable more precise modeling and better adaptation to changes in demand time series.

Growing interest in more sophisticated models has also led to the application of graph-based architectures [59,60] and Transformers (an architecture based on self-attention mechanisms that dynamically weights relationships among sequence time points, capturing short- and long-range dependencies) [61–64], as evidenced in [65], where a graph network with temporal convolutions (GLFN-TC) was introduced to capture spatial and temporal relationships simultaneously. This model outperformed LSTM, CNN-LSTM, and Informer, achieving better results across multiple international datasets. These approaches represent a new frontier in the use of DL for energy forecasting by enabling full power grids to be modeled as dynamic graphs where system structure and evolution are learned simultaneously.

For *Transformers*, Ref. [63] developed a model called Romember, a Transformer architecture that incorporates rotary positional encoding (RoPE), designed to overcome standard-model limitations in representing relative position in long sequences. Applied to Estonia's electrical system, Romember achieved a MAPE of 3.19%, substantially improving over RNN, GRU, and baseline Transformers, and showing strong stability for use in smart grids under energy transition. On the other hand, Ref. [62] evaluated the Temporal Fusion Transformer (TFT) in a hierarchical network of 70 substations in Germany (2019–2021), finding that although it does not always outperform LSTM in daily forecasting, it shows notable

advantages in weekly horizons and hierarchical configurations (MAPE between 2.43% and 3.76%). This reinforces its usefulness for operational planning with aggregated structures.

Despite their high potential, *Deep Learning* models present significant challenges. Training is usually intensive in data and computational resources, requires careful hyperparameter selection, and has limited interpretability, which can hinder adoption by system operators who need clear explanations for decision-making. However, recent literature has shown that these limitations can be mitigated through techniques such as transfer learning, proper regularization, and the integration of explanatory layers or attention mechanisms.

Ultimately, *Deep Learning* models represent the state of the art in electricity demand forecasting. Their ability to capture complex relationships, integrate multiple data types, and adapt to nonlinear patterns makes them key tools for addressing modern power-system challenges. Although their application requires suitable technical infrastructure and specialized knowledge, their adoption is rapidly expanding worldwide, especially in systems where precision and resilience are priorities.

Despite their strong predictive performance, *Deep Learning* models often involve high computational costs and limited interpretability. These limitations have motivated the development of hybrid approaches, discussed in Section 8, which aim to combine the interpretability and structural advantages of statistical methods with the representational capacity of neural network architectures.

The models reviewed in this section can be grouped according to their main characteristics, advantages, and limitations, as summarized in Table 4, which presents a detailed comparative summary of the analyzed Deep Learning models. As visual support, Figure 10 summarizes the taxonomy of Deep Learning architectures used in demand forecasting, showing their organization from dense networks to recurrent, convolutional, and attention-based models.

Table 4. Description of Deep Learning models for electricity demand forecasting.

Model	Typical Application	Key Advantages	Limitations
<i>Artificial Neural Networks (ANN)</i>	Multivariable demand prediction with dense-layer architectures.	Can model nonlinear relationships; adaptable to many contexts.	Does not leverage temporal relationships; requires extensive calibration; prone to overfitting.
<i>Recurrent Neural Networks (RNN)</i>	Modeling short sequences or series with immediate dependence among values.	Captures short-term memory in sequences; flexible structure.	Vanishing/exploding gradient problems; poor performance on long sequences.
<i>Long Short-Term Memory (LSTM)</i>	Hourly demand prediction in the presence of long-term dependencies.	Excellent long-term memory capacity; highly accurate for high-resolution series.	Requires large datasets and high computational power; long training time.
<i>Gated Recurrent Units (GRU)</i>	Sequential forecasting where lower complexity than LSTM is required with similar performance.	More compact and faster structure; good generalization; efficient in the short term.	May lose accuracy relative to LSTM on long sequences.
<i>Convolutional Neural Networks (CNN)</i>	Detection of repetitive patterns in smart-meter data or minute-level granular demand.	Highly efficient; captures spatial or local relationships in time series.	Does not retain long-term temporal memory unless combined with recurrent models.
<i>Temporal Convolutional Networks (TCN)</i>	Alternative to RNN for long sequences through causal convolution.	Captures temporal dependencies without recurrence; better parallelization.	Less widespread; complex hyperparameter tuning; requires long temporal padding.
<i>Transformers</i>	Modeling long series with high cross-dependence and multiple variables.	Captures contextual attention; high accuracy in multivariate environments.	High computational consumption; still less common in energy applications.

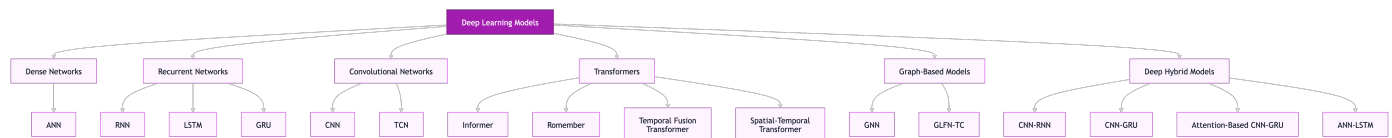


Figure 10. Taxonomy of Deep Learning models for electricity demand forecasting.

8. Hybrid Models

Hybrid models integrate different methodological approaches to leverage the complementary strengths of each one. These combinations may involve the union of statistical models with *Machine Learning* or *Deep Learning* techniques, as well as the incorporation of mathematical transformation or decomposition methods such as Empirical Mode Decomposition (EMD), Wavelet Transform, fuzzy logic, or seasonal decomposition. In this sense, the hybrid approach is not limited only to combining models of different nature, but can also consist of the sequential application of adaptive preprocessing techniques that extract structural components of the time series. This strategy allows both global patterns and local dynamics at different temporal scales to be captured, improving the system's predictive capability. However, integrating multiple stages may increase model complexity, raise the number of parameters to estimate, and generate risks associated with overfitting, information redundancy, and error propagation across stages.

In electricity demand forecasting, the search for models that balance accuracy, robustness, interpretability, and computational efficiency has led to the development of hybrid approaches, which combine elements from different statistical, *Machine Learning*, and *Deep Learning* methodologies. These models are based on the recognition that no individual approach is optimal in all contexts or under all input conditions, so intelligent integration of techniques can provide substantial improvements in predictive capability and adaptability of forecasting systems.

Hybrid models can be classified into two broad categories: those that combine statistical models with machine-learning models and those that integrate statistical models with deep-learning models. In both cases, the motivation is the same: to compensate for the weaknesses of one method with the strengths of another, whether by combining interpretability and efficiency of statistical models with the ability of AI-based models to capture nonlinearities, or by integrating diverse architectures to cover different temporal and spatial scales or behavior patterns.

One of the most representative hybrid approaches is the combination of traditional time-series models with neural networks. In ref. [66], an ARIMA–LSTM model was developed for short-term forecasting in China, achieving a MAPE of 2.83%, significantly lower than ARIMA (6.96%) or LSTM (2.86%) used individually. This model is based on a sequential logic in which the ARIMA component captures the linear structure of the series, while the LSTM network models residual nonlinear components. This type of combined approach is particularly useful when the series has marked seasonality and trends, but also abrupt variations that cannot be explained by traditional linear models.

Similarly, Ref. [67] presents a combination of statistical models with neural networks; in this case, a hybrid model combining Empirical Mode Decomposition (EMD) with an LSTM recurrent neural network. The model uses EMD to decompose the time series into intrinsic mode functions (IMFs), and these IMFs are used as input to LSTM, allowing the model to better capture nonlinear and non-stationary characteristics of the time series. This model was compared with other Deep Learning models such as RNN and LSTM. Results show that the EMD+LSTM model outperforms the others in forecasting, with errors between 5 and 8%, standing out in peak-pressure and seasonal-variation prediction. This integration notably improves forecast accuracy and stability. Although EMD-based hybrids

can improve the decomposition of non-stationary demand series, the procedure is not without limitations. In practice, EMD may exhibit mode mixing, end effects, and sensitivity to local noise; therefore, the decomposition stage should be interpreted as a preprocessing aid rather than as a guaranteed representation of the underlying signal. Furthermore, the effectiveness of decomposition-based approaches depends on the quality of the input data, since noise, anomalies, and measurement inconsistencies may propagate through the decomposition process and ultimately affect downstream forecasting performance.

Other studies have explored the use of Prophet as a statistical component to capture seasonality and trend, complemented by networks such as LSTM or GRU to model residuals. For example, in ref. [68], an optimized Prophet–LSTM with a BPNN network was applied to data from Belgium, achieving a MAPE of 0.91%, while in ref. [69], Prophet and GRU were integrated for forecasting in industrial contexts, notably improving performance relative to their individual components.

On the other hand, hybrid models have also shown advantages in contexts with incomplete data or complex time series. In ref. [70], cubic *splines* interpolation was used as an imputation technique together with a tree-based ensemble model, achieving a MAPE of 0.648% even with 30% missing data.

However, building hybrid models also entails important challenges. The main difficulty lies in correct integration of components, both at the technical level (model interfaces, input/output synchronization) and at the theoretical level (avoiding redundancies, ensuring complementarity). Likewise, hybrid models tend to be more difficult to interpret in general without sufficient technical knowledge, which can limit applicability in environments where full traceability is required for regulatory or auditing purposes. In many cases, the accuracy improvement obtained with hybrid models must be carefully weighed against increased complexity, computational cost, and maintenance.

Ultimately, hybrid models represent a logical and powerful evolution within the electricity-demand forecasting ecosystem. Their ability to combine the best of different worlds (the robustness and explainability of statistical models, the flexibility of ML models, and the representational power of DL) makes them highly valuable tools, especially in modern electrical systems characterized by high uncertainty, multiple exogenous factors, and growing precision requirements. Although their design, implementation, and validation are not trivial tasks, the results reported in the literature show that, when properly applied, hybrid models not only outperform individual models, but also open new possibilities for developing smarter, adaptive, and more reliable forecasting systems.

The models reviewed in this section can be grouped according to their main characteristics, advantages, and limitations, as summarized in Table 5, which presents a detailed comparative summary of the analyzed hybrid models. Complementarily, Figure 11 illustrates the taxonomy of hybrid models, highlighting the most representative combinations among statistical, Machine Learning, and Deep Learning components.

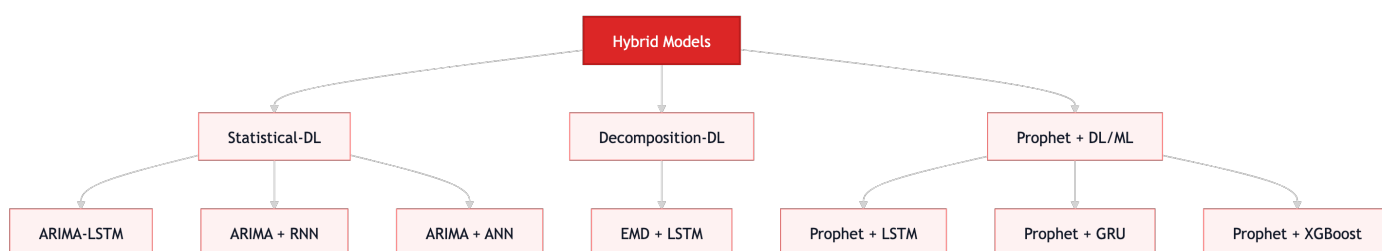


Figure 11. Taxonomy of hybrid models for electricity demand forecasting.

Table 5. Description of hybrid models for electricity demand forecasting.

Model	Typical Application	Key Advantages	Limitations
ARIMA–LSTM	Short-term forecasting in series with mixed structure.	Captures both linear and nonlinear relationships; improves accuracy in demand peaks.	Requires careful calibration; risk of redundancy between components.
EMD + LSTM	Non-stationary series with multiple frequencies.	Improves decomposition and modeling of complex fluctuations.	High sensitivity to preprocessing; greater technical complexity.
Prophet + LSTM/GRU	Demand with a strong cyclical component and presence of exogenous events.	Interpretable; handles holidays and climatic variables; adaptable to short and medium term.	Higher computational load; requires hyperparameter tuning.
Prophet + XGBoost	Very short-term forecasting with enriched data.	Good handling of external variables and demand peaks.	Requires feature engineering; depends on data quality.
ARIMA + RNN/ANN	Multicomponent forecasting in the presence of significant noise.	Allows assigning an optimal model to each component (trend, cycle, residual).	Complex integration; high design and implementation effort.

9. Discussion

9.1. Qualitative Analysis

The evolution of electricity demand forecasting methods has produced a growing diversity of approaches ranging from traditional statistical models to sophisticated deep-learning architectures and hybrid combinations. This diversity has significantly enriched the field by providing tools with different levels of accuracy, adaptability, and technical complexity, but it has also introduced new challenges when selecting the most appropriate model for each operational context. Therefore, any model comparison must go beyond evaluations based only on point error metrics such as MAE or MAPE, and adopt a broader perspective that includes robustness under adverse conditions, computational cost, interpretability, scalability, hierarchical forecast coherence, and adaptive capacity.

Grouping models through a structured taxonomy is essential to guide both research and practice in electricity demand forecasting. The diversity of available approaches (from classical statistical techniques to state-of-the-art modular architectures) can make it difficult to identify the most suitable method for a specific operational setting. A taxonomy that classifies models by methodological family, application horizon, and required data type reduces this complexity, facilitates objective comparison, and provides a useful reference framework for researchers and energy-sector professionals. In addition, this classification makes explicit the relationships and complementarities among model families, which is crucial for designing hybrid architectures and identifying methodological gaps that can motivate future research lines. Without a clear taxonomy, the risk of selecting models arbitrarily or duplicating effort in already existing solutions increases considerably, compromising design-process efficiency and forecast quality.

In contexts where time series are stable, show clear seasonality, and have low sensitivity to nonlinear exogenous variables, classical statistical models such as ARIMA or multiple linear regression remain relevant and effective tools. Their main virtue is high interpretability, allowing clear understanding of each variable's specific contribution to projected demand, which is especially valuable for long-term energy planning in regulated environments [1,21,28]. These models have achieved errors below 5% in medium- and

long-term forecasts when weather and calendar variables are properly integrated. However, their performance tends to deteriorate significantly in the presence of nonlinearities, high volatility, or cross-interactions among multiple exogenous factors, as often occurs in densely connected urban grids or contexts with high penetration of intermittent renewable energy.

In this sense, defining a model's applicability to the demand forecasting problem cannot be reduced to a single selection criterion. The reviewed evidence shows that the most decisive aspects include, first, the forecast horizon, since relevant dynamics and the type of patterns the model must capture vary substantially across short-, medium-, and long-term settings. Second, data nature and availability directly condition the viability of each approach: Deep Learning models require large volumes of high-resolution data, whereas statistical models can operate with shorter and less granular series. Third, the model's ability to integrate exogenous variables (meteorological, socioeconomic, calendar-based, or atypical-event related) is critical in systems where demand responds to multiple simultaneous factors. These criteria are complemented by the interpretability required by the regulatory environment, the computational cost associated with training and inference, and model robustness under noisy, incomplete, or structurally shifting data. It is the joint weighting of these factors, rather than absolute superiority in any isolated metric, that determines model suitability for a specific electricity-demand forecasting problem.

Given these limitations, *Machine Learning* models have gained ground, with algorithms such as *Random Forest* (RF), *Support Vector Regression* (SVR), and XGBoost standing out. These models have shown superior capability to capture complex relationships in multidimensional, nonlinear, and highly dynamic environments, without requiring prior functional specification. In short- and very-short-term scenarios in particular, they have reached errors below 2%, in some cases even outperforming deep neural-network-based models [27,41,43]. Random Forest has proven especially effective by offering an attractive balance among accuracy, tuning-process stability, and partial interpretability, delivering robust performance across multiple datasets [25]. XGBoost, in turn, has stood out in peak-demand prediction thanks to its ability to incorporate adaptive loss functions, which translates into greater sensitivity to extreme events and high temporal granularity [43].

However, the true methodological revolution has come with the widespread adoption of *Deep Learning* models, led by architectures such as LSTM, GRU, and CNN. These models have established new accuracy standards, with MAPE below 1.5% in hourly and daily forecasting, even under highly seasonal conditions or strong variability [10,50–52,60]. Their ability to learn complex hierarchical representations directly from raw data, without explicit modeling, makes them especially powerful in systems where electricity demand responds to multiple temporal scales and nonlinearities. However, this predictive power comes with high computational cost, strong sensitivity to parametrization, and limited internal traceability of decisions, which may constrain adoption in critical or highly regulated systems.

To mitigate these limitations, hybrid models have emerged as a strategic alternative that combines the best of each approach. Their value does not lie only in an additive combination of techniques, but in a functional decomposition of the problem that addresses different series components (trend, seasonality, residual noise) with specialized methods. In particular, approaches integrating models such as Prophet, time-frequency decomposition techniques, and deep neural networks have shown substantial improvements in both accuracy and robustness [68,69,71,72]. These models have proven effective not only for short-term peak forecasting, but also for capturing structural variations in medium-term scenarios, particularly in urban environments where the signal is contaminated by multiple operating modes. Hybrid and ensemble models are among the most consistent alternatives in contexts where data include a broad set of exogenous variables and exhibit complex dynamics [71].

From the perspective of the presented taxonomy, each model family makes a differentiated and complementary contribution to solving the forecasting problem. Statistical models constitute the analytical foundation of the field: their ability to decompose the series into interpretable components (trend, seasonality, and residuals) makes them irreplaceable tools for structural understanding of demand and for establishing baselines against which more complex approaches can be evaluated. Their main contribution lies in prediction transparency and traceability, which are essential attributes in regulated contexts where decisions must be auditable. *Machine Learning* models, in turn, significantly expand the solution space by enabling the capture of nonlinear relationships and integration of multiple information sources without prior functional specification. Their contribution is especially visible in high-dimensional short- and medium-term scenarios, where statistical models show structural limitations. Within this family, ensemble models (such as Random Forest and XGBoost) have shown a particularly valuable balance among accuracy, stability, and computational efficiency. Regarding *Deep Learning*, its main contribution lies in the ability to model long-range temporal dependencies and extract hierarchical representations directly from high-resolution data, positioning these models as state of the art in short-term forecasting accuracy when abundant information is available. Finally, hybrid models contribute the ability to functionally decompose the problem, assigning the most suitable method to each series component, thus overcoming the individual limitations of each family and providing more robust and adaptable forecasts for complex operational environments.

An emerging approach that deserves special attention in this comparative landscape is the *Mixture of Experts* (MoE) paradigm, which offers a highly promising alternative for addressing the growing complexity of energy forecasting. Unlike traditional models (where a single architecture attempts to capture all time-series complexity), MoE models rely on a modular structure in which multiple experts are trained to specialize in different regions of the input space or different behavioral patterns. An additional network, called the gating network, dynamically assigns weights to each expert according to the current input features, enabling adaptive and localized combination of predictions [46–48,73]. This approach presents several advantages that make it particularly attractive. First, it enables significant accuracy improvements by specializing each expert in a subset of the problem, such as specific day segments, particular weather conditions, or user types. This problem partitioning allows each submodel to reach higher representational capacity without increasing overall system complexity. Second, MoE offers a path toward better interpretability: by analyzing which experts activate under specific scenarios, it becomes possible to understand what patterns the system recognizes, facilitating adoption in regulated or operationally transparency-sensitive environments. In addition, from a computational viewpoint, MoE can be efficient, since typically only a subset of experts is activated per instance, significantly reducing inference-time computational load. This makes it an attractive option for real-time applications or resource-constrained systems.

From a structural perspective, the superiority of Mixture of Experts (MoE) over traditional ensemble approaches such as *stacking* and *bagging* arises from the way predictions are combined. In conventional ensemble methods, the combination weights are either fixed, as in the uniform averaging strategy commonly used in *bagging*, or globally learned through a meta-learner trained on validation data. As a result, the same combination rule is applied to all inputs during inference, regardless of the characteristics of the forecasting scenario. In contrast, MoE architectures employ a *gating network* that dynamically assigns weights to different experts according to the characteristics of the current input. This mechanism allows the model to activate specialized experts depending on the operating conditions of the system. Consequently, MoE models can adapt their predictive behavior to different demand regimes, such as extreme weather conditions, holidays, or normal operational periods,

which is not inherently possible in static ensemble formulations. The effectiveness of this adaptive mechanism has been empirically demonstrated in recent studies. In ref. [46], MoE achieved superior performance compared to LSTM and Temporal Convolutional Networks (TCN) under extreme operating conditions, reaching MAPE values between 3.1% and 3.9%, while maintaining competitive accuracy during normal conditions with a MAPE of 2.9%. Similarly, Ref. [47] reported that combining MoE with Federated Learning produced improvements of up to 41% in *zero-shot* forecasting scenarios. This represents a particularly challenging generalization setting in which static ensemble methods usually exhibit poor performance due to the absence of retraining signals [48]. This input-conditional behavior constitutes a fundamental qualitative difference between MoE and classical ensemble methods: whereas Bagging assigns uniform weights regardless of the input and Stacking applies fixed weights learned during training, MoE continuously reassigns the contribution of each expert based on the current input context, effectively acting as a context-aware coordinator rather than a static aggregator. In electricity demand forecasting, this translates into the ability to selectively activate specialist models under different operating regimes (such as extreme weather events, holidays, or structural demand shifts) without relying on a single combination rule across all scenarios.

This adaptive routing mechanism, however, also introduces practical challenges for real-world deployment. One of the most important is *expert collapse*, a training instability in which the gating network converges to activating only a small subset of experts, thereby reducing the benefits of expert specialization [74]. This issue is commonly mitigated through auxiliary load-balancing losses during training [47]. In addition, the gating network can be sensitive to delays in real-time data availability. In operational environments such as dispatch control centers or smart-grid monitoring systems, sensor readings and demand measurements do not always arrive synchronously [46]. Because expert routing depends on the current input features, latency or missing information may lead to incorrect expert assignment, degrading forecasting performance precisely during the operational transitions where adaptive specialization is expected to provide the greatest benefit.

Another crucial aspect in model comparison is forecast nature: point or probabilistic. Far from being optional or aesthetic, probabilistic forecasting has become an essential requirement in risk-sensitive applications, such as reserve management, electricity-market operation, or contractual planning. Tools such as *Quantile Regression Forests* (QRF), probabilistic *stacking* techniques, and *Mixture of Experts* (MoE)-type models enable not only estimation of well-defined confidence intervals, but also adaptation of models to different operational-risk levels [20,46–48]. Evaluation of these models should include specific metrics such as CRPS (Continuous Ranked Probability Score), quantiles, and the Winkler index. These metrics allow model performance to be assessed according to risk associated with decisions based on the forecast, facilitating integration into real operating environments.

Forecast quality also depends critically on two cross-cutting operational factors. The first is weather-input quality: many academic evaluations use ex-post weather, i.e., real climate data observed after the prediction horizon, which can artificially overestimate model accuracy. As noted in [19], studies should clearly report the source and status of the weather forecast used, as well as the temporal-validation protocol, to ensure comparability of results. Designing more rigorous evaluation protocols requires, at minimum, adopting walk-forward validation schemes that preserve the temporal ordering of the data, using forecast inputs that are realistically available at decision time rather than ex-post observed values, and reporting accuracy in a disaggregated manner across operating conditions (such as weekdays versus weekends, seasonal periods, or anomalous events), rather than relying solely on average error metrics that may conceal systematic failures under specific conditions. The second factor is concept drift, i.e., structural changes in demand patterns

due to social, technological, or economic transformations [31]. In this case, resilient approaches have been proposed, such as online ensembles with RNN and ARIMA that update weights according to recent performance, or transfer-learning strategies to stabilize errors in low-history contexts [31,75]. These strategies have shown that the best model is not a static choice, but a continuous process of updating, drift monitoring, and expert reweighting. When drift is gradual and no explicit change-point signal is available, dynamic combination of multiple models is preferable to rolling updates of a single model. A diverse set of models can progressively absorb structural changes through adaptive weighting, whereas a single model depends on external triggers to initiate retraining and may degrade in the interim without explicit detection mechanisms [31]. The need for continuous adaptation has also motivated growing interest in continual-learning approaches, which seek to update forecasting models as new information becomes available while retaining previously acquired knowledge [76]. Such methods are particularly relevant under concept drift, where consumption patterns evolve gradually and static models may become progressively less representative of the underlying system. However, continual adaptation introduces an additional challenge known as catastrophic forgetting, whereby learning new patterns can degrade performance on previously learned regimes [77]. Consequently, future forecasting systems must balance adaptability and knowledge retention, combining drift-aware updating mechanisms with the ability to preserve useful historical information.

Hierarchical coherence is another increasingly relevant aspect in complex systems. This principle states that forecasts must remain consistent across different aggregation levels (for example, feeder-level forecasts should sum to substation-level forecasts, and in turn to the national total). Hierarchical probabilistic methods that use forecast reconciliation, such as approaches based on *Aggregated Model Integration* (AMI) or *Quantile Regression Forests* (QRF), adjust local and global forecasts while guaranteeing coherent relationships throughout the hierarchy, substantially improving forecast usefulness for coordinated operational planning [36]. This methodological line aligns with the most recent industrial recommendations, which emphasize the need for reconciliation protocols, smart weather-station selection, and multiscale verification as critical links in the forecasting chain [20].

Finally, the adoption of advanced architectures such as *Transformers*, *Graph Neural Networks* (GNN), and *Temporal Convolutional Networks* (TCN) represents a promising frontier, especially for modeling complex spatiotemporal dependencies. Although these architectures offer incremental accuracy gains, their high training cost, sensitivity to tuning, and lack of standardization in temporal-validation protocols require careful implementation decisions [53–55,57]. In practice, they are often used as components within hybrid-model pipelines that maximize accuracy without fully sacrificing traceability. Cases such as CNN-GRU for very-short- and short-term forecasting [42], or Prophet-XGBoost-CatBoost combinations [78], illustrate this principle of balance between performance and operational viability.

It is worth noting that the business value of forecasting is not limited to technical metrics. Recent studies have shown that models with similar average errors may differ significantly in economic impact. For example, XGBoost has simultaneously achieved MAPE of 1.41% and effective minimization of financial imbalance cost using metrics such as CLFBM (Cost of Load Forecasting Based Mistakes) [43]. This evidence suggests that model evaluation should integrate, in addition to accuracy and risk, a cost-benefit dimension that reflects real operational impact.

From an operational perspective, computational cost varies significantly across forecasting models. Classical statistical approaches such as SARIMA stand out for low memory and processing requirements in both training and inference, making them suitable for resource-constrained environments. By contrast, *Machine Learning* techniques such as

Random Forest and XGBoost involve an intermediate burden: although training may become costly due to hyperparameter optimization, they remain CPU-viable with moderately efficient inference. On the other hand, *Deep Learning* models (LSTM, GRU, CNN, and *Transformers*) have high computational cost, especially during training, due to intensive resource use (backpropagation, gradient storage, etc.). In particular, self-attention-based Transformers exceed traditional RNNs in complexity. This increase also translates into significant energy cost: recent studies report consumption of up to 1 MWh per training run in complex, non-optimized models [79]. In the energy domain, although DL offers notable accuracy gains (from 20% up to 40% according to [29]), this implies a computational-sustainability penalty. Alternatives such as GRU (which can match LSTM performance with lower load) or efficient hybrid models emerge as strategic solutions, confirming that architectural efficiency can outperform simple complexity growth [80]. These considerations become even more relevant when forecasting systems must operate under deployment constraints. While large Deep Learning and hybrid architectures can achieve excellent predictive performance, their computational requirements may limit their practical adoption in resource-constrained environments. Consequently, future research should not only pursue higher forecasting accuracy, but also consider model efficiency, scalability, and operational feasibility as complementary design objectives.

In conclusion, comparison among electricity demand forecasting models should not be understood as a competition among isolated approaches, but as a search for integrative, adaptive, and cost-effective architectures. Convergence among statistical, ML, and DL techniques, systematic incorporation of probabilistic approaches, and adoption of robust practices around granularity and hierarchy point toward a new generation of forecasting systems designed for complex, dynamic, and data-driven decision-making environments.

To complement the discussion presented in this section, Tables 6–9 were structured to synthesize the information discussed in this analysis of electricity-demand forecasting techniques and methods. These tables provide a detailed taxonomy of the addressed models, classifying them by methodological family in Table 6, application horizon in Table 7, and required data type in Table 8. Table 9 is also included, allowing cross-visualization of relationships among these three key criteria. This systematization offers a valuable tool for both researchers and energy-sector professionals, facilitating contextualized selection of the most appropriate technique or method according to the characteristics of the analyzed electrical system.

Table 6. Taxonomy by model family for electricity demand forecasting.

Family	Model	Brief Description	Main Advantages	Limitations	Refs.
Statistical	ARIMA/SARIMA/SARIMAX	Autoregressive models that capture trend and seasonality.	High interpretability and low computational cost.	Limited capability under nonlinearities or multiple variables (especially in ARIMA and SARIMA).	[1,13,14,32,33]
Statistical	Multivariable regression	Models linear relationships between demand and exogenous variables (weather, GDP, population).	Simple, transparent, and explainable.	Limited accuracy in dynamic or complex systems.	[21,23,28,30]

Table 6. Cont.

Family	Model	Brief Description	Main Advantages	Limitations	Refs.
Machine Learning	Support Vector Regression (SVR)	Maximum-margin algorithm that fits nonlinear hyperplanes.	High accuracy and robustness to noise.	Requires intensive calibration and can be costly to train.	[12,27,38,39]
Machine Learning	Bagging (Random Forest)	Decision-tree ensemble with random sampling.	Robust to overfitting and good generalization capability.	Low interpretability and dependence on number of trees.	[22,25,26,40,41]
Machine Learning	Boosting (XG-Boost/AdaBoost)	Sequential ensemble that optimizes residual errors.	Excellent accuracy and adaptability.	High training cost and overfitting risk.	[20,22,42,43]
Machine Learning	Stacking	Combination of several models for final prediction or intervals.	Improves accuracy and uncertainty calibration; robust under variability.	High complexity; difficult interpretation; requires model diversity.	[20]
Machine Learning	Mixture of Experts (MoE)	Set of specialized experts and gating network.	Modularity, interpretability, and inference efficiency.	Theoretical complexity and high design cost; susceptible to expert collapse during training; sensitive to data latency in real-time routing.	[46–48,73]
Deep Learning	ANN	Networks with fully connected dense layers.	Model complex nonlinear relationships.	Require large data volumes and extensive tuning.	[6,49,58]
Deep Learning	RNN/LSTM/GRU	Networks with temporal memory for sequences.	Capture long-term dependencies; ideal for time series.	High computational cost and limited interpretability.	[10,50–52]
Deep Learning	CNN/TCN	One-dimensional convolutions for local-pattern extraction.	Highly efficient on high-frequency data.	Limited for long dependencies without additional components.	[17,54,55]
Deep Learning	Transformers	Models with self-attention mechanisms that capture long-range dependencies.	High accuracy and efficient parallelization.	Require large data volumes and intensive training.	[61,62,64,81]
Deep Learning	GNN	Neural networks on graphs (nodes and edges).	Capture spatial and topological correlations.	Complex to define and train; high computational demand.	[59,60]
Deep Learning	Deep hybrids	Combine deep architectures to leverage different capabilities.	High predictive capacity and adaptability.	Require significant resources and fine optimization.	[40,54,55,57]
Hybrids	ARIMA–LSTM/EMD + LSTM/Prophet + LSTM/GRU	Integrate linear models with neural networks or ML.	Combine interpretability with high accuracy.	Calibration and integration difficulty.	[66–70]

Table 7. Taxonomy by time horizon for electricity demand forecasting.

Horizon	Typical Scale	Applications	Most Used Models	Advantages	Refs.
Short term	Minutes to days	Hourly control, demand response, daily operation.	LSTM, GRU, TCN, CNN, RF, ARIMA, XGBoost, <i>Transformers</i> , MoE.	High accuracy and immediate response.	[10,17,41,47,48,50,54,57]
Medium term	Weeks to months	Operational planning, maintenance, and tariff design.	SVR, ANN, LSTM, ensemble models, CNN–GRU.	Good balance between accuracy and computational cost.	[27,28,30,35,42,58,82]
Long term	Years or more	Grid expansion planning and energy-policy design.	Multivariable regression, GNN, ANN.	Captures macroeconomic and structural effects.	[21,23,45,46]

Table 8. Taxonomy by data types used for electricity demand forecasting.

Data Type	Description	Examples	Compatible Models	Advantages/Applications	Refs.
Temporal univariate	Only the historical demand series.	Historical kWh.	ARIMA, ETS, LSTM, GRU.	Simple, lightweight, and low computational cost.	[1,13,14]
Temporal multivariate	Include exogenous variables in addition to demand.	Weather, calendar, GDP, prices.	ANN, SVR, CNN, LSTM, <i>Transformers</i> , MoE.	Greater contextual accuracy and explanatory capability.	[10,42,46,50,54,57]
Spatial/geographic	Data structured by region, grid, or electrical nodes.	Substations, nodes.	GNN, CNN, MoE.	Capture spatial correlations and topological patterns.	[48,53,60]
High frequency/IoT	Data recorded in seconds or minutes.	Smart meters, SCADA.	CNN, TCN, DL hybrids.	Enable granular, real-time forecasting.	[42,52,54]
Socioeconomic data	Macroeconomic and behavioral variables.	GDP, trade, tariffs.	Multivariable regression, ensemble models.	Useful for medium- and long-term scenarios.	[21,23,45,50]

Table 9. Summary of electricity demand forecasting models by integrated taxonomy.

Family	Model	Horizon	Data	Key Advantages	Refs.
Statistical	ARIMA/SARIMA/SARIMAX	Short/medium	Univariate	Simplicity, explainability, and efficient execution.	[1,13,14,32,33]
Statistical	Multivariable regression	Medium/long	Multivariate; socioeconomic	Transparency and usefulness in planning scenarios.	[21,23,28,30]
<i>Machine Learning</i>	Support Vector Regression (SVR)	Short/medium	Multivariate; socioeconomic	Captures nonlinearities and generalizes well with limited data.	[12,27,38,39]
<i>Machine Learning</i>	Bagging (Random Forest)	Short/medium/long	Multivariate; spatial	High accuracy and resistance to overfitting.	[25,26,40,41]
<i>Machine Learning</i>	Boosting (XG-Boost/AdaBoost)	Short/medium	Multivariate; economic	Excellent predictive performance and good scalability.	[20,22,42,43]

Table 9. Cont.

Family	Model	Horizon	Data	Key Advantages	Refs.
Machine Learning	Stacking	Short	Univariate or mixed	Improves accuracy and uncertainty estimation; high complexity.	[20]
Machine Learning	Mixture of Experts (MoE)	Short	Multivariate	Modularity, adaptability, and inference efficiency.	[46–48,73]
Deep Learning	RNN/LSTM/GRU	Short/medium	Multivariate; high-frequency	Model short- and long-term temporal dependencies.	[10,50–52]
Deep Learning	CNN/TCN	Short	High-frequency; spatial	Efficiently extract local and periodic patterns.	[17,54,55]
Deep Learning	Transformers	Short/medium	Multivariate; long sequences	Capture long-range dependencies with efficient parallelization.	[61,62,64,81]
Deep Learning	GNN	Medium/long	Spatial; structural	Model electric networks and complex spatial correlations.	[59,60]
Deep Learning	Deep hybrids	Short/medium	Multivariate; high-frequency	Synergy among deep architectures to improve accuracy and robustness.	[40,54,55,57]
Hybrids	ARIMA–LSTM/EMD + LSTM/Prophet + LSTM/GRU	Short/medium	Multivariate	Combine linear and nonlinear patterns in an integrated scheme.	[66–70]

9.2. Quantitative Analysis

The discussion presented so far has made it possible to identify the strengths and limitations of each model family from a qualitative perspective. However, to guide model selection more rigorously according to the operational context, this analysis must be complemented with a quantitative assessment that synthesizes, on a common scale, the relative performance of each family against clearly defined evaluation criteria. For this purpose, this subsection introduces a rating scale from 1 to 5 (where 5 indicates superior performance and 1 indicates low performance), based on the criteria defined in the methodology: *forecast horizon, performance metrics, modeling capability and robustness, adaptability and scalability, interpretability, computational cost, and handling of exogenous variables*. The scores are reported in Tables 10 and 11 and visualized in Figure 12.

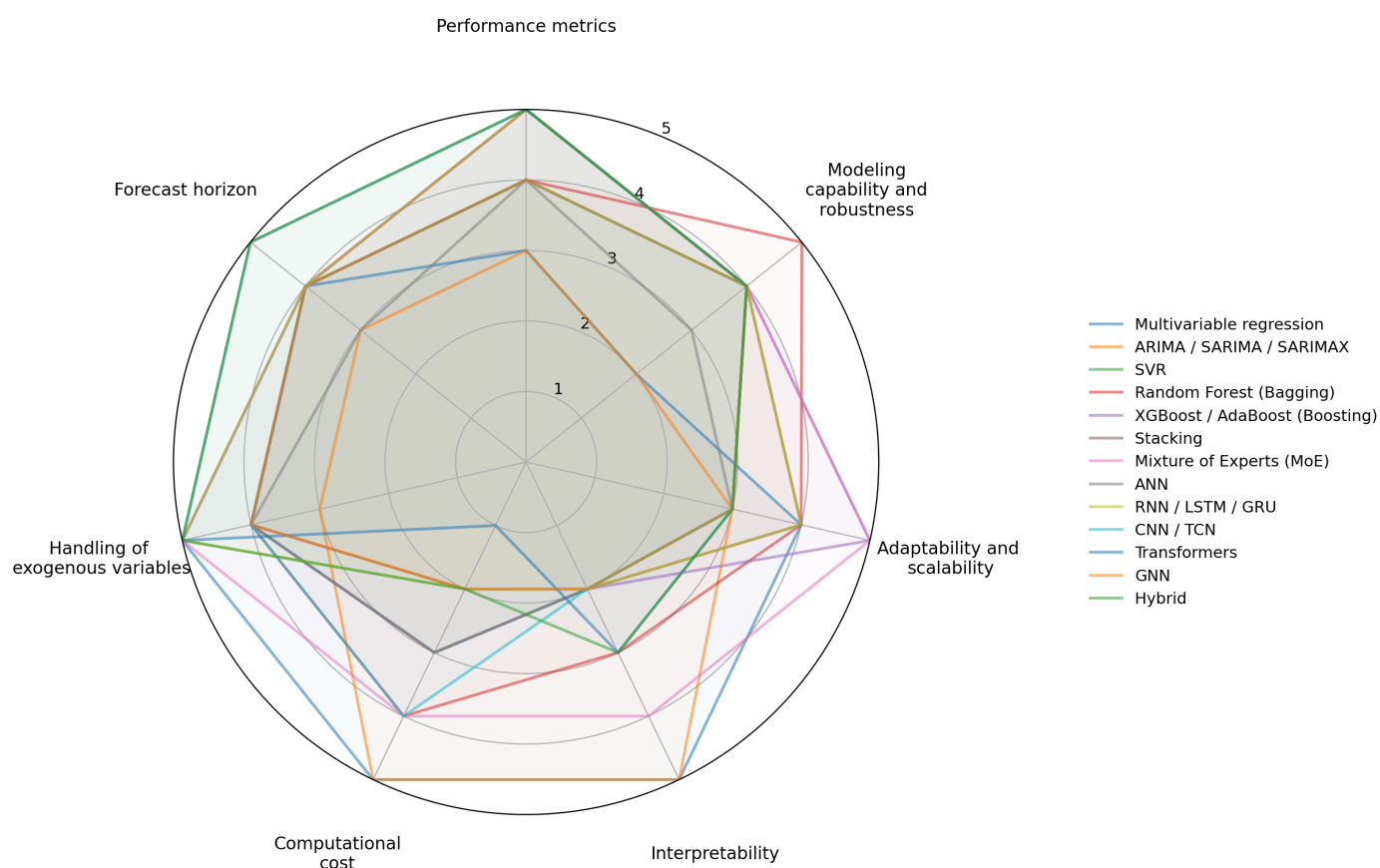
Table 10. Scores of electricity demand forecasting models by evaluation criteria (scale 1–5, where 5 = best).

Family/Model	Performance Metrics	Modeling Capability	Adaptability and Scalability	Interp.	Computational Cost	Exogenous Variables	Horizon Adequacy
ARIMA/SARIMA/SARIMAX	3	2	3	5	5	3	3
Multivariable regression	3	2	4	5	5	5	4
SVR	4	4	3	2	3	4	4
Random Forest (RF)	4	5	4	3	4	4	4
XGBoost/AdaBoost	5	4	5	2	3	4	4
Stacking/ensembles	5	4	4	2	2	4	4
MoE (<i>Mixture of Experts</i>)	5	4	5	4	4	5	4
ANN/MLP	4	3	3	2	3	4	3
RNN/LSTM/GRU	5	4	4	2	2	5	4
CNN/TCN	4	4	3	2	4	4	4
Transformers	5	4	3	3	1	5	5
GNN	4	4	3	2	2	4	4
Hybrids	5	4	3	3	2	5	5

Table 11. Ranking of electricity demand forecasting models by average score.

Ranking	Model	Average
1	MoE (<i>Mixture of Experts</i>)	4.43
2	Multivariable regression	4.00
2	Random Forest (RF)	4.00
3	XGBoost/AdaBoost	3.86
3	Hybrids	3.86
4	Transformers	3.71
4	RNN/LSTM/GRU	3.71
5	Stacking/ensembles	3.57
5	CNN/TCN	3.57
6	ARIMA/SARIMA/SARIMAX	3.43
6	SVR	3.43
7	GNN	3.29
8	ANN/MLP	3.14

Radar chart

**Figure 12.** Radar chart of scores for representative models, based on Table 10.

The scoring used in Table 10 is intended as an evidence-based comparative synthesis rather than a formal decision-making model. Because the reviewed studies differ substantially in terms of datasets, forecasting horizons, validation protocols, and the availability of exogenous variables, the application of formal Multi-Criteria Decision-Making methods would introduce a degree of numerical precision that is not fully supported by the available evidence. For this reason, the score anchors were established a priori and applied consistently across all model families to reduce arbitrariness and enhance transparency.

The evaluation criteria and the meaning of the scores are presented below.

- **Forecast horizon:** Evaluates how suitable the approach is for short-, medium-, or long-term operation. A score of 1 indicates that the model is only applicable to a very specific horizon and with low performance; 2 reflects limited applicability to a single horizon with acceptable performance; 3 corresponds to competitiveness in one or two horizons (e.g., short/medium term); 4 implies good performance across multiple horizons with minor adjustments; and 5 reflects excellent performance across several forecasting horizons.
- **Performance metrics:** Summarizes the relative performance reported in standard metrics (MAE, RMSE, MAPE) and forecast accuracy. A score of 1 corresponds to high errors that fail to accurately capture demand patterns and behavior (e.g., MAPE above 15%); 2 indicates partial improvement but systematic failures in peaks, valleys, or transitions (MAPE 10–15%); 3 corresponds to a competitive baseline with acceptable performance under normal conditions, although it deteriorates in complex scenarios (MAPE 5–10%); 4 implies low and relatively stable errors across multiple scenarios and horizons (MAPE 3–5%); and 5 denotes superior and consistent performance, effectively capturing peaks, seasonality, and extreme events (MAPE below 3% consistently across different datasets and operating conditions). For studies that do not report MAPE, the assessment is based on equivalent evidence from other commonly reported metrics, such as MAE, RMSE, or R^2 .
- **Modeling capability and robustness:** Measures the ability to capture relationships (including nonlinearities) and maintain performance under noise, missing data, outliers, and regime changes. A score of 1 implies severe performance degradation under any minor disturbance (noise, missing data, or isolated outliers); 2 reflects significant sensitivity to moderate noise or outliers, with unstable performance across periods; 3 denotes stability under normal operating conditions but noticeable deterioration under moderate disturbances such as atypical seasonal changes; 4 implies tolerance to noise, missing data, outliers, and moderate shifts with limited loss of predictive power; and 5 is observed in models that remain stable under abrupt regime changes, extreme weather events, or incomplete data.
- **Adaptability and scalability:** Considers how easily the model can be adjusted to new contexts (periods/regions/conditions) and how it behaves as the number of series, granularity, dimensionality, and data volume increase. A score of 1 implies that the model does not scale and is difficult to adapt; 2 reflects poor scalability and adaptability only through intensive intervention (feature engineering or retraining); 3 corresponds to moderate scalability and reasonable adaptability with retraining and moderate adjustments; 4 is assigned to models that scale well with parallelization and standard hardware; and 5 corresponds to highly scalable models that can expand to new series, regions, or variables with minimal pipeline changes and approximately proportional cost growth.
- **Interpretability:** Evaluates forecast traceability and explainability, meaning the ability to identify the contribution of each variable, rule, or component to the forecast. A score of 1 corresponds to a black-box model with no useful operational explanation; 2 implies limited explainability dependent on auxiliary techniques (SHAP, LIME), without guaranteed sufficient traceability; 3 indicates partial interpretability, where internal model indicators such as feature importance or attention weights are available, but without full decision-process traceability; 4 denotes high interpretability with directly analyzable model components (e.g., domain-specialized experts); and 5 is assigned to models with full interpretability, with coefficients, rules, or components (trend, seasonality, residual) that are directly readable and auditable.

- **Computational cost:** Quantifies the resources required for training, updating, and inference (time, memory, energy), as well as viability for production or real-time use. A score of 1 reflects very high costs that make deployment impractical without specialized infrastructure; 2 implies high cost that limits frequent retraining and requires dedicated infrastructure; 3 corresponds to moderate cost and feasibility with standard resources (CPU/GPU); 4 is assigned to models whose training is feasible on CPU within reasonable times and whose inference is efficient for frequent or near real-time operation; and 5 corresponds to models with lightweight training and inference, executable on minimal hardware and suitable for continuous real-time updating.
- **Handling of exogenous variables:** Assesses the type and number of exogenous variables that can be effectively incorporated to improve forecasts. A score of 1 indicates that the model is purely univariate with no ability to incorporate external information; 2 reflects a limited ability to include few exogenous variables in a rigid way (e.g., only through specific transformations or feature engineering); 3 corresponds to moderate integration of exogenous variables that notably improves model performance; 4 denotes flexible integration of multiple exogenous variable types with the ability to capture cross-interactions; and 5 is assigned to models that dynamically and adaptively integrate multiple heterogeneous exogenous sources, with substantial forecast improvement.

The scores in Table 10 were assigned by analyzing the evidence discussed throughout the document under the previously defined methodology. The profile of each model family is justified below through an analysis of the established criteria. Based on the average scores across the evaluated criteria, the electricity demand forecasting models were ranked, as shown in Table 11. The results reveal that the Mixture of Experts model achieves the highest overall score, suggesting that it is a highly promising approach for electricity demand forecasting.

Models in the ARIMA family (SARIMA, SARIMAX) constitute the classical statistical baseline for electricity demand forecasting. In terms of performance metrics (3), they have shown low errors under favorable conditions, such as MAPE of 0.87% [1] or 1.73% [37]; however, error increases rapidly under complex dynamics, as the same approach rose from 4.13% in short-term forecasting to 55% in medium-term forecasting [1]. Modeling capability and robustness (2) reflect their structural limitation: by assuming linearity, they do not capture nonlinearities or cross-interactions and deteriorate under abrupt events [15] or regime changes [13]. Adaptability and scalability (3) are moderate: these models generalize well if seasonal patterns are preserved, but degrade under *concept drift* [31]; they scale by training one model per series, although per-series parametrization is operationally costly. Interpretability (5) and computational cost (5) are their strongest advantages, since their structure (AR/MA terms, explicit seasonality, exogenous coefficients in SARIMAX) provides full traceability [4,12], and training is feasible on limited CPU hardware, even in real time as shown by SARIMA-ONS [33]. Handling exogenous variables (3) is limited in ARIMA/SARIMA (univariate), but extended in SARIMAX, where adding temperature, humidity, and calendar variables improved accuracy by 22% in [32]. Horizon adequacy (3) reflects relevance for short- and medium-term forecasting in stable series [13,14].

Multivariable regression shares ARIMA's strengths in interpretability (5) and computational cost (5), but differs in its native ability to incorporate exogenous variables. In performance metrics (3), it achieves good results when the demand–exogenous relationship is linear, as in [21] with MAPE of 2.10%, outperforming XGBoost, LightGBM, and LSTM, and showing competitiveness against ML/DL in [30]. However, under nonlinear relationships, the error increases substantially. Modeling capability and robustness (2) reflect restrictions to linear relations and sensitivity to multicollinearity [23,28]. Adapt-

ability and scalability (4) are higher than ARIMA, scaling linearly with variables and enabling regional hierarchization [23], although they suffer under abrupt changes [31]. Handling exogenous variables (5) is the strongest among statistical models, since regression directly quantifies the influence of heterogeneous predictors (weather, GDP, HDD/CDD, calendar) [21,28]. Horizon adequacy (4) reflects utility for medium- and long-term forecasting with macroeconomic variables [23], and, to a lesser extent, competitiveness in short-term settings.

SVR models represent the first step toward nonlinear modeling. Performance metrics (4) are supported by results such as MAPE of 1.35% with metaheuristic optimization [38], errors of 0.56–1.03% with SVR-RBFNN [39], and superiority over k -NN and RF in [27]. Although several studies report low MAPE values, these results are obtained under highly optimized configurations, including metaheuristic parameter optimization or hybrid initialization with extended Kalman filtering, and therefore are not consistently representative of standard SVR implementations across diverse evaluation settings [38,39]. Modeling capability and robustness (4) are explained by kernel-based projections into high-dimensional spaces, which capture nonlinearities and tolerate noise [38,39]. Adaptability and scalability (3) are moderate: these models generalize well with appropriate regularization [12], but their algorithmic complexity limits scalability for large volumes. Interpretability (2) is low when nonlinear kernels are used. Computational cost (3) reflects reasonable inference but expensive training for large datasets [12]. Handling exogenous variables (4) is flexible, integrating weather and calendar variables with implicit interaction capture [27,39]. Horizon adequacy (4) is favorable for short- and medium-term forecasting [27].

Random Forest is positioned as one of the most balanced approaches. Performance metrics (4) include MAPE of 0.161% and R^2 of 0.979 in ultra-short-term forecasting [41], as well as superiority over statistical and deep models [25]; it does not reach 5 because *boosting* tends to outperform it in scenarios with highly complex interactions [40]. Modeling capability and robustness (5) are high: the *bagging* ensemble strategy, by processing random subsets, provides resistance to overfitting, tolerance to heterogeneous variables, and demonstrated stability across multiple regions [25,26]. Adaptability and scalability (4) benefit from inherent parallelism and the ability to handle many variables [41], with successful application from ultra-short-term to long-term forecasting [45], although performance may deteriorate under *concept drift* [31]. Interpretability (3) comes from metrics such as built-in feature importance, but without unique global rules. Computational cost (4) is favorable, since training is CPU-feasible and parallelizable [25]. Handling exogenous variables (4) natively integrates different variable types [40]. Horizon adequacy (4) spans from granular short-term forecasting [41] to long-term forecasting [45].

Boosting methods rank among the highest in terms of accuracy and adaptability–scalability. Performance metrics (5) include MAPE of 1.41% with XGBoost [43] and 0.72% with AdaBoost, with an 82% reduction versus individual models [44]. Modeling capability and robustness (4) are high because they capture complex nonlinearities, although without adequate regularization they may overfit [42]. Adaptability and scalability (5) stem from optimized implementations that exploit parallelism and *sparsity-aware learning* [43]. Interpretability (2) is limited without methods such as SHAP or LIME. Computational cost (3) reflects expensive training due to hyperparameter search. Handling exogenous variables (4) integrates weather, economic, and calendar variables [43]. Horizon adequacy (4) reflects excellent performance in short- and medium-term forecasting [22,43].

Stacking aims to maximize accuracy by combining complementary models. Performance metrics (5) are evidenced by probabilistic *stacking* in [20] and MAPE of 0.648% with 30% missing data in [70]. Modeling capability and robustness (4) are high when

base models are diverse [20]. Adaptability and scalability (4) are good if automation and parallelization are applied, although higher complexity increases overfitting risk. Interpretability (2) is limited when heterogeneous architectures are combined. Computational cost (2) is high because multiple base models plus a meta-model must be trained. Handling exogenous variables (4) inherits the capability of the strongest base model. Horizon adequacy (4) positions these methods well in short- and medium-term forecasting [20].

MoE models obtain the most balanced profile. Performance metrics (5) include MAPE of 2.9% on regular days and 1.79% on holidays, outperforming LSTM and TCN [46], improvements of 6–9% on static data and up to 41% in *zero-shot* settings [47], and MSE reductions of 22% with TriForecaster [48]. Modeling capability and robustness (4) arise from conditional activation of condition-specialized experts (heat waves, holidays, normal operation), enabling dynamic adaptation to regime changes [46]. Adaptability and scalability (5) are maximal due to modularity, which allows adding experts without retraining the full system [73]; additionally, Federated Learning reduced training time by 18% [47]. Interpretability (4) is supported because expert-activation analysis can reveal which patterns the system recognizes [46]. Computational cost (4) is favorable since only a subset of experts is activated per instance [46]. Handling exogenous variables (5) integrates heterogeneous information through both the experts and the *gating network* [46,48,73]. Horizon adequacy (4) reflects strong performance in operational short-term forecasting and multi-condition/multi-region scenarios [46–48].

Dense-architecture ANN models represent the entry point to Deep Learning. Performance metrics (4) are supported by their universal approximation capability, as they outperformed multiple regression and fuzzy logic for horizons up to 168 h [6], and improved when including climate variables in regions with high thermal sensitivity [49]. Modeling capability and robustness (3) reflect that they capture moderate nonlinearities but are sensitive to overfitting without proper regularization [49]. Adaptability and scalability (3) are moderate, depending on data volume. Interpretability (2) is limited (black-box behavior). Computational cost (3) is intermediate with fast inference. Handling exogenous variables (4) is flexible, integrating multiple numerical and categorical inputs [6,49]. Horizon adequacy (3) reflects utility in short- and medium-term forecasting, but lower temporal specialization than RNN or TCN [50].

Recurrent networks (RNN/LSTM/GRU) represent the natural framework for modeling temporal dependencies. Performance metrics (5) include daily MAPE of 1.01% and weekly MAPE of 1.09% [50], 1.5% hourly error while integrating weather and renewables [51], and more than 20% reduction versus SVR and Gradient Boosting using PowerLSTM [52]. Modeling capability and robustness (4) are explained by gating mechanisms that capture temporal dependencies, although they remain sensitive to parametrization [16]. Adaptability and scalability (4) reflect good generalization with sufficient data, but susceptibility to *concept drift* [75] and memory requirements for multi-series operation. Interpretability (2) is limited without attention mechanisms. Computational cost (2) is high due to backpropagation through time. Handling exogenous variables (5) is among the strongest in DL: these models natively integrate heterogeneous inputs in their sequential flow, including volatility weighting with MIDAS [16] and smart-meter data [52]. Horizon adequacy (4) reflects systematic excellence in short-term forecasting [10,50,51].

Convolutional networks (CNN) offer a homogeneous profile that balances predictive power and efficiency. Performance metrics (4) include R^2 of 0.992 with a simple CNN [17], performance comparable to LSTM with lower cost in TCN [54], and improvements in probabilistic intervals with TCN-QR and TCN-MC [55]. Modeling capability and robustness (4) benefit from causal convolutions and progressive dilation, remaining stable without gradient vanishing. Adaptability and scalability (3) are moderate and depend on temporal

window design [54]. Interpretability (2) is low. Computational cost (4) is an advantage, as these models can be more parallelizable than RNNs with efficient real-time inference [54]. Handling exogenous variables (4) integrates multiple input channels through multivariate convolutions. Horizon adequacy (4) reflects utility in short-term, high-granularity forecasting [54,55].

Transformers represent the most advanced frontier in multivariate modeling. Performance metrics (5) include MAPE of 3.19% with Romember [63] and 2.43–3.76% with TFT in hierarchical configurations [62]. Modeling capability and robustness (4) are high thanks to multi-scale self-attention, but sensitive to *tuning* [62]. Adaptability and scalability (3) reflect that parallelism improves scalability compared with RNNs, but computational complexity and memory demand limit long sequences. Interpretability (3) offers partial visualization of attention weights. Computational cost (1) is the main drawback, since training is extremely expensive. Handling exogenous variables (5) is excellent because known and observed temporal variables can be explicitly integrated [62]. Horizon adequacy (5) reflects optimal adaptation for both short and long sequences, as well as hierarchical scenarios [62,63].

Graph neural networks (GNN) introduce the topological structure of power grids. Performance metrics (4) are evidenced by GLFN-TC outperforming LSTM, CNN-LSTM, and Informer [65], along with additional improvements reported in [59,60]. Modeling capability and robustness (4) capture spatial correlations, but require an appropriate graph representation [60]. Adaptability and scalability (3) are moderate: deployment is feasible when topology is stable, but constrained by graph size and memory. Interpretability (2) requires auxiliary analysis methods. Computational cost (2) is high due to graph propagation. Handling exogenous variables (4) integrates node attributes such as historical demand and local climate [59]. Horizon adequacy (4) reflects utility with spatial coherence in short- and medium-term forecasting [59,60].

Hybrid models functionally decompose the series, assigning each component to the most suitable method. Performance metrics (5) include ARIMA–LSTM with MAPE of 2.83% versus 6.96% for standalone ARIMA [66], Prophet–LSTM with 0.91% [68], and EMD+LSTM improving peak prediction by 5–8% [67]. Modeling capability and robustness (4) benefit from the layered pipeline structure, where statistical components such as ARIMA, Prophet, or EMD first isolate linear, trend, seasonal, or low-frequency patterns before the Deep Learning component models the remaining nonlinear dynamics. This complementary decomposition improves robustness to moderate noise and demand variability [66,68]. However, multi-stage architectures are not free from limitations. Errors introduced during decomposition or preprocessing stages may propagate through subsequent learning components, creating cascading error propagation that can reduce robustness under pronounced structural changes or inaccurate intermediate representations [68,69]. Adaptability and scalability (3) are moderate: the complexity of the multi-level *pipeline* increases error-propagation risk. Interpretability (3) improves over pure DL due to the partial transparency of the statistical component [68]. Computational cost (2) is high due to multiple sequential stages. Handling exogenous variables (5) integrates variables through the statistical component (especially Prophet) and captures nonlinear interactions through the DL component [68–70]. Horizon adequacy (5) reflects versatility in short- and medium-term forecasting [68,69].

10. Conclusions

The bibliometric analysis presented in Section 4 yields three main findings that guide the interpretation of the rest of this article. First, the field’s thematic structure confirms the consolidation of three core pillars: classical statistical methods, ensemble-based *Machine Learning* models, and recurrent/convolutional *Deep Learning* architectures, each with differ-

entiated research communities and technical vocabularies. Second, the temporal evolution reveals a migration from classical artificial neural networks toward more sophisticated models (LSTM, CNN, TCN), and, more recently, a growing interest in ensemble methods and hybrid models that integrate statistical and artificial intelligence components. Third, the fragmentation of co-authorship networks indicates opportunities for interdisciplinary integration that could catalyze the development of new architectures, such as *Mixture of Experts* models, which require the convergence of knowledge from different communities.

The detailed study of the diverse approaches used for electricity demand forecasting clearly shows both the technological evolution of the field and the persistent challenges that forecasting systems still face. From classical statistical approaches, which provide highly interpretable models but remain limited when confronted with complex phenomena, to modern *Deep Learning* architectures and hybrid models, the literature reflects a constant tension among accuracy, complexity, transparency, and adaptability.

One of the main conclusions from this analysis is that selecting the optimal model is not a purely technical decision, but one that must be deeply grounded in the operational, regulatory, and technological context in which the forecasting system will be implemented. Although models such as ARIMA or multivariable regression continue to show notable performance in stable and well-structured contexts, their effectiveness is rapidly surpassed by machine-learning models when series exhibit high nonlinearity, complex cross-interactions, or significant exogenous perturbations.

On the other hand, *Machine Learning* models have proven particularly effective at capturing hidden patterns and complex dynamics, although their applicability at industrial scale requires careful consideration regarding feature selection, robust performance evaluation, and operational implementation in mission-critical environments. Similarly, *Deep Learning* models have established new standards in predictive accuracy, especially in systems with high temporal resolution or large volumes of multivariate data. However, their high computational cost, the need for specialized infrastructure, and limited interpretability create important barriers to adoption in some electricity-market segments, particularly those with budget constraints or strict regulatory requirements.

In this scenario, hybrid models emerge as an attractive intermediate solution by combining the robustness of traditional techniques with the flexibility of modern methods. Their ability to simultaneously capture linear and nonlinear components, or to integrate different information sources and types, makes them ideal candidates for complex and highly dynamic environments. Nevertheless, their implementation requires a systemic perspective and a high technical level, so their use should be carefully evaluated according to available resources and capabilities.

Given the diversity of approaches and analyzed results, it is necessary to recognize that there is no universally superior model for electricity demand forecasting. Although some algorithms have shown better performance in specific contexts (for example, LSTM networks in hourly horizons or hybrid models in the presence of multiple exogenous variables), this does not imply a general rule applicable to every scenario, nor is it possible to establish a normative method that directly and definitively associates each time horizon (short, medium, or long term) with a specific model family. The applicability of an approach fundamentally depends on data quality, availability, and granularity, as well as on the degree of knowledge about system structure and the factors that determine its behavior. For this reason, designing a robust forecasting system must begin with a rigorous exploratory analysis of available data to identify relevant patterns and understand their underlying dynamics. Only from this diagnosis is it possible to select and calibrate the model that best fits the particular problem conditions, integrating expert knowledge with the predictive capabilities of different techniques.

One of the most promising directions in this field is the exploration of *Mixture of Experts* (MoE)-type models, which could represent a new paradigm in energy-demand forecasting. Their modular structure, which assigns specialized experts to different input types, enables not only accuracy improvements but also greater interpretability, scalability, and computational efficiency. This architecture could be key for developing adaptive models capable of responding quickly to structural changes or unforeseen events, without fully retraining the entire system. In particular, their ability to integrate heterogeneous experts—for example, statistical models, neural networks, and tree-based techniques—under a coordinated co-operation logic positions them as one of the most complete and adaptable alternatives for future challenges in the electricity sector.

Looking ahead, research directions should focus on building integrative methods that not only improve accuracy in diverse environments, but also guarantee traceability, robustness, and operational feasibility of developed models. This includes advancing interpretability techniques for complex models, developing continuous-training schemes for dynamic environments, exploring efficient low-latency algorithms for real-time execution, and establishing broader metrics that include not only prediction errors but also economic impacts, resilience to structural data shifts, and adaptability to new data sources.

In summary, electricity-demand forecasting cannot be viewed solely as a strictly mathematical or technical problem, but as a systemic challenge requiring cross-cutting integration of statistical, computational, energy, and organizational knowledge. The current state of the art has shown that a wide range of tools is available, each with strengths and limitations, and that model selection must be guided by contextualized and strategic criteria. Ultimately, real progress in the field will depend not only on model sophistication, but on our collective ability to integrate models efficiently, transparently, and responsibly into the decision-making processes that support modern power-system operation.

Author Contributions: Conceptualization, A.S.T.-S., Á.J.-D. and W.M.V.-A.; methodology, A.S.T.-S., Á.J.-D. and W.M.V.-A.; investigation, A.S.T.-S.; writing—original draft preparation, A.S.T.-S.; writing—review and editing, Á.J.-D. and W.M.V.-A.; supervision, Á.J.-D. and W.M.V.-A. All authors have read and agreed to the published version of the manuscript.

Funding: The authors gratefully acknowledge the financial support provided by Universidad de Antioquia (UdeA) and the Colombian Ministry of Science, Technology, and Innovation (MinCiencias) through the Patrimonio Autónomo Fondo Nacional de Financiamiento para la Ciencia, la Tecnología y la Innovación, Francisco José de Caldas (Perseo Alliance, Contract No. 112721-392-2023).

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article. The public data sources used for the descriptive figures are cited in the manuscript.

Acknowledgments: The authors acknowledge the Research Group GIMEL.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
CNN	Convolutional Neural Network
CDD	Cooling Degree Days
CRPS	Continuous Ranked Probability Score
EMD	Empirical Mode Decomposition

ETS	Exponential Smoothing (state space)
GNN	Graph Neural Network
GRU	Gated Recurrent Units
HDD	Heating Degree Days
LSTM	Long Short-Term Memory
LTLF	Long-Term Load Forecasting
MAPE	Mean Absolute Percentage Error
MIDAS	Mixed Data Sampling
MLP	Multi-Layer Perceptron
MoE	Mixture of Experts
MTLF	Medium-Term Load Forecasting
QRF	Quantile Regression Forests
RBFNN	Radial Basis Function Neural Network
RNN	Recurrent Neural Network
RF	Random Forest
SARIMA	Seasonal Autoregressive Integrated Moving Average
SARIMAX	Seasonal ARIMA with eXogenous variables
sMAPE	Symmetric Mean Absolute Percentage Error
STLF	Short-Term Load Forecasting
SVM	Support Vector Machines
SVR	Support Vector Regression
TCN	Temporal Convolutional Network
TFT	Temporal Fusion Transformer
XGBoost	Extreme Gradient Boosting

References

1. Medina, D.O.G.; García, G.A.M. Estimación del consumo eléctrico colombiano en el corto y largo plazo empleando regresión multivariable y series temporales. *Av. Investig. Ing.* **2017**, *14*, 155–168. <https://doi.org/10.18041/1794-4953/avances.1.1294>.
2. Ullah, K.; Ahsan, M.; Hasanat, S.M.; Haris, M.; Yousaf, H.; Raza, S.F.; Tandon, R.; Abid, S.; Ullah, Z. Short-Term Load Forecasting: A Comprehensive Review and Simulation Study With CNN-LSTM Hybrids Approach. *IEEE Access* **2024**, *12*, 111858–111881. <https://doi.org/10.1109/ACCESS.2024.3440631>.
3. Myagmarjav, T.; Bandi, T.; Bayar, B.E.; Ohtsuka, S. Changes in Electric Power Demands of Mongolian and Japanese Local Areas in 2020 and Discussion on Seasonal Electricity Usage Based on the Temperature Dependence. In *Proceedings of the 2022 9th International Conference on Condition Monitoring and Diagnosis (CMD)*; IEEE: Piscataway, NJ, USA, 2022; pp. 355–359. <https://doi.org/10.23919/CMD54214.2022.9991573>.
4. Raza, M.Q.; Khosravi, A. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. *Renew. Sustain. Energy Rev.* **2015**, *50*, 1352–1372. <https://doi.org/10.1016/j.rser.2015.04.065>.
5. Enerdata. World Energy Statistics | Enerdata. Available online: <https://yearbook.enerdata.net/> (accessed on 12 October 2025).
6. Al-ani, B.R.K.; Erkan, T.E. A Study of Load Demand Forecasting Models in Electricity using Artificial Neural Networks and Fuzzy Logic Model. *Int. J. Eng.* **2022**, *35*, 1111–1118. <https://doi.org/10.5829/IJE.2022.35.06C.02>.
7. Zahin, S.; Latif, H.; Paul, S.; Azeem, A. A Comparative Analysis of Power Demand Forecasting with Artificial Intelligence and Traditional Approach. *Int. J. Bus. Inf. Syst.* **2013**, *13*, 359–380. <https://doi.org/10.1504/IJBIS.2013.054469>.
8. Pourdayaei, A.; Mohammadi, M.; Mubarak, H.; Abdellatif, A.; Karimi, M.; Gryazina, E.; Terzija, V. A new framework for electricity price forecasting via multi-head self-attention and CNN-based techniques in the competitive electricity market. *Expert Syst. Appl.* **2024**, *235*, 121207. <https://doi.org/10.1016/j.eswa.2023.121207>.
9. Saglam, M.; Spataru, C.; Karaman, O.A. Electricity Demand Forecasting with Use of Artificial Intelligence: The Case of Gokceada Island. *Energies* **2022**, *15*, 5950. <https://doi.org/10.3390/en15165950>.
10. Gil, R.; Ignacio, A. Hourly Electricity Consumption Forecasting for Antioquia-Colombia Using Statistical-Machine Learning Models. Master's Thesis, Universidad Nacional de Colombia, Bogota, Colombia, 2024.
11. Zor, K.; Timur, O.; Teke, A. A state-of-the-art review of artificial intelligence techniques for short-term electric load forecasting. In *Proceedings of the 2017 6th International Youth Conference on Energy (IYCE)*; IEEE: Piscataway, NJ, USA, 2017; pp. 1–7. <https://doi.org/10.1109/IYCE.2017.8003734>.

12. Hernandez, L.; Baladron, C.; Aguiar, J.M.; Carro, B.; Sanchez-Esguevillas, A.J.; Lloret, J.; Massana, J. A Survey on Electric Power Demand Forecasting: Future Trends in Smart Grids, Microgrids and Smart Buildings. *IEEE Commun. Surv. Tutor.* **2014**, *16*, 1460–1495. <https://doi.org/10.1109/SURV.2014.032014.00094>.
13. Shah, I.; Iftikhar, H.; Ali, S.; Wang, D. Short-Term Electricity Demand Forecasting Using Components Estimation Technique. *Energies* **2019**, *12*, 2532. <https://doi.org/10.3390/en12132532>.
14. Shah, I.; Iftikhar, H.; Ali, S. Modeling and Forecasting Medium-Term Electricity Consumption Using Component Estimation Technique. *Forecasting* **2020**, *2*, 163–179. <https://doi.org/10.3390/forecast2020009>.
15. Buechler, E.; Powell, S.; Sun, T.; Zanolco, C.; Astier, N.; Bolorinos, J.; Flora, J.; Boudet, H.; Rajagopal, R. Power and the Pandemic: Exploring Global Changes in Electricity Demand During COVID-19. *arXiv* **2020**, <https://doi.org/10.48550/arXiv.2008.06988>.
16. Choi, E.; Cho, S.; Kim, D.K. Power Demand Forecasting using Long Short-Term Memory (LSTM) Deep-Learning Model for Monitoring Energy Sustainability. *Sustainability* **2020**, *12*, 1109. <https://doi.org/10.3390/su12031109>.
17. Kang, T.; Lim, D.Y.; Tayara, H.; Chong, K.T. Forecasting of Power Demands Using Deep Learning. *Appl. Sci.* **2020**, *10*, 7241. <https://doi.org/10.3390/app10207241>.
18. Tan, M.; Yuan, S.; Li, S.; Su, Y.; Li, H.; He, F. Ultra-Short-Term Industrial Power Demand Forecasting Using LSTM Based Hybrid Ensemble Learning. *IEEE Trans. Power Syst.* **2020**, *35*, 2937–2948. <https://doi.org/10.1109/TPWRS.2019.2963109>.
19. Misiurek, K.; Olkusi, T.; Zyśk, J. Review of Methods and Models for Forecasting Electricity Consumption. *Energies* **2025**, *18*, 4032. <https://doi.org/10.3390/en18154032>.
20. Dudek, G. Stacking for Probabilistic Short-term Load Forecasting. *arXiv* **2024**. <https://doi.org/10.48550/arXiv.2406.10718>.
21. Hong, J.H.; Lee, G.C. Monthly Load Forecasting in a Region Experiencing Demand Growth: A Case Study of Texas. *Energies* **2025**, *18*, 4135. <https://doi.org/10.3390/en18154135>.
22. Mathew, A.; Chikte, R.; Sadanandan, S.K.; Abdelaziz, S.; Ijaz, S.; Ghaoud, T. Medium-term feeder load forecasting and boosting peak accuracy prediction using the PWP-XGBoost model. *Electr. Power Syst. Res.* **2024**, *237*, 111051. <https://doi.org/10.1016/j.epsr.2024.111051>.
23. Adeoye, O.; Spataru, C. Modelling and forecasting hourly electricity demand in West African countries. *Appl. Energy* **2019**, *242*, 311–333. <https://doi.org/10.1016/j.apenergy.2019.03.057>.
24. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *PLoS Med.* **2021**, *18*, e1003583. <https://doi.org/10.1371/journal.pmed.1003583>.
25. Dudek, G. A Comprehensive Study of Random Forest for Short-Term Load Forecasting. *Energies* **2022**, *15*, 7547. <https://doi.org/10.3390/en15207547>.
26. Dudek, G. Short-Term Load Forecasting Using Random Forests. In *Intelligent Systems '2014*; Springer: Cham, Switzerland, 2014. https://doi.org/10.1007/978-3-319-11310-4_71.
27. Jawad, M.; Nadeem, M.S.A.; Shim, S.O.; Khan, I.R.; Shaheen, A.; Habib, N.; Hussain, L.; Aziz, W. Machine Learning Based Cost Effective Electricity Load Forecasting Model Using Correlated Meteorological Parameters. *IEEE Access* **2020**, *8*, 146847–146864. <https://doi.org/10.1109/ACCESS.2020.3014086>.
28. Mirasgedis, S.; Sarafidis, Y.; Georgopoulou, E.; Lalas, D.P.; Moschovits, M.; Karagiannis, F.; Papakonstantinou, D. Models for mid-term electricity demand forecasting incorporating weather influences. *Energy* **2006**, *31*, 208–227. <https://doi.org/10.1016/j.energy.2005.02.016>.
29. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V.; Semenoglou, A.A.; Mulder, G.; Nikolopoulos, K. Statistical, machine learning and deep learning forecasting methods: Comparisons and ways forward. *J. Oper. Res. Soc.* **2023**, *74*, 840–859. <https://doi.org/10.1080/01605682.2022.2118629>.
30. Lee, G.C. A Regression-Based Method for Monthly Electric Load Forecasting in South Korea. *Energies* **2024**, *17*, 5860. <https://doi.org/10.3390/en17235860>.
31. Jagait, R.K.; Fekri, M.N.; Grolinger, K.; Mir, S. Load Forecasting Under Concept Drift: Online Ensemble Learning with Recurrent Neural Network and ARIMA. *IEEE Access* **2021**, *9*, 98992–99008. <https://doi.org/10.1109/ACCESS.2021.3095420>.
32. Elamin, N.; Fukushige, M. Modeling and forecasting hourly electricity demand by SARIMAX with interactions. *Energy* **2018**, *165*, 257–268. <https://doi.org/10.1016/j.energy.2018.09.157>.
33. Anh, N.T.N.; Anh, N.N.; Thang, T.N.; Solanki, V.K.; Crespo, R.G.; Dat, N.Q. Online SARIMA applied for short-term electricity load forecasting. *Appl. Intell.* **2023**, *54*, 1003–1019. <https://doi.org/10.1007/s10489-023-05230-y>.
34. Mustafa, A.T.; Al-Yozbaky, O.S.A.D. Forecasting Energy Demand and Generation Using Time Series Models: A Comparative Analysis of Classical, Grey, Fuzzy, and Intelligent Approaches. *Frankl. Open* **2025**, *12*, 100350. <https://doi.org/10.1016/j.fraope.2025.100350>.
35. Zimmermann, M.; Ziel, F. Efficient mid-term forecasting of hourly electricity load using generalized additive models. *Appl. Energy* **2025**, *388*, 125444. <https://doi.org/10.1016/j.apenergy.2025.125444>.

36. Taieb, S.B.; Taylor, J.W.; Hyndman, R.J. Hierarchical Probabilistic Forecasting of Electricity Demand With Smart Meter Data. *J. Am. Stat. Assoc.* **2021**, *116*, 27–43. <https://doi.org/10.1080/01621459.2020.1736081>.
37. Melchor Cruz, E.A.; Hernández Armenta, L.A.; Badaoui, M. Seasonal forecasting of peak electricity demand using spectral analysis. *Results Eng.* **2025**, *28*, 107338. <https://doi.org/10.1016/j.rineng.2025.107338>.
38. Hu, Z.; Bao, Y.; Xiong, T. Electricity Load Forecasting Using Support Vector Regression with Memetic Algorithms. *Sci. World J.* **2013**, *2013*, 292575. <https://doi.org/10.1155/2013/292575>.
39. Ko, C.N.; Lee, C.M. Short-term load forecasting using SVR (support vector regression)-based radial basis function neural network with dual extended Kalman filter. *Energy* **2013**, *49*, 413–422. <https://doi.org/10.1016/j.energy.2012.11.015>.
40. Kayacı Çodur, M. Ensemble Machine Learning Approaches for Prediction of Türkiye's Energy Demand. *Energies* **2024**, *17*, 74. <https://doi.org/10.3390/en17010074>.
41. Gao, J.; Wang, K.; Kang, X.; Li, H.; Chen, S. Ultra-short-term electricity load forecasting based on improved random forest algorithm. *AIP Adv.* **2023**, *13*, 065208. <https://doi.org/10.1063/5.0153550>.
42. Sajjad, M.; Khan, Z.A.; Ullah, A.; Hussain, T.; Ullah, W.; Lee, M.Y.; Baik, S.W. A Novel CNN-GRU-Based Hybrid Approach for Short-Term Residential Load Forecasting. *IEEE Access* **2020**, *8*, 143759–143768. <https://doi.org/10.1109/ACCESS.2020.3009537>.
43. Beloev, H.I.; Saitov, S.R.; Filimonova, A.A.; Chichirova, N.D.; Babikov, O.E.; Iliev, I.K. Short-Term Electrical Load Forecasting Based on XGBoost Model. *Energies* **2025**, *18*, 5144. <https://doi.org/10.3390/en18195144>.
44. Gao, T.; Niu, D.; Ji, Z.; Sun, L. Mid-term electricity demand forecasting using improved variational mode decomposition and extreme learning machine optimized by sparrow search algorithm. *Energy* **2022**, *261*, 125328. <https://doi.org/10.1016/j.energy.2022.125328>.
45. Gulaydin, O.; Mourshed, M. Machine learning for subnational residential electricity demand forecasting to 2050 under shared socioeconomic pathways: Comparing tree-based, neural and kernel methods. *Energy* **2025**, *336*, 138195. <https://doi.org/10.1016/j.energy.2025.138195>.
46. Liu, Y.; Zhang, K.; Chen, S.; Qian, X.; Zhou, Y.; Zhan, X.; Chen, J. Interpretable prediction of short-term load curve for extreme scenarios based on mixture of experts model. *Electr. Power Syst. Res.* **2025**, *247*, 111730. <https://doi.org/10.1016/j.epsr.2025.111730>.
47. Sievers, J.; Blank, T.; Simon, F. Advancing Accuracy in Energy Forecasting using Mixture-of-Experts and Federated Learning. In *Proceedings of the Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems, e-Energy '24*; ACM: New York, NY, USA, 2024; pp. 65–83. <https://doi.org/10.1145/3632775.3661945>.
48. Zhu, Z.; Zeng, Z.; Chen, Q.; Yang, L.; Liu, P.; Chen, W.; Sun, L. TriForecaster: A Mixture of Experts Framework for Multi-Region Electric Load Forecasting with Tri-dimensional Specialization. *arXiv* **2025**. <https://doi.org/10.48550/arXiv.2508.09753>.
49. Baranitharan, B.; Sivapragasam, C.; Rajesh, K. Energy Demand Prediction Considering Weather Change Effects using Artificial Neural Network. In *Proceedings of the 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*; IEEE: Piscataway, NJ, USA, 2021; pp. 916–921. <https://doi.org/10.1109/ICAC3N53548.2021.9725452>.
50. ul Islam, B.; Ahmed, S.F. Short-Term Electrical Load Demand Forecasting Based on LSTM and RNN Deep Neural Networks. *Math. Probl. Eng.* **2022**, *2022*, 2316474. <https://doi.org/10.1155/2022/2316474>.
51. Waheed, W.; Xu, Q.; Aurangzeb, M.; Iqbal, S.; Dar, S.H.; Elbarbary, Z.M.S. Empowering data-driven load forecasting by leveraging long short-term memory recurrent neural networks. *Heliyon* **2024**, *10*, e40934. <https://doi.org/10.1016/j.heliyon.2024.e40934>.
52. Cheng, Y.; Xu, C.; Mashima, D.; Thing, V.L.L.; Wu, Y. PowerLSTM: Power Demand Forecasting Using Long Short-Term Memory Neural Network. In *Proceedings of the Advanced Data Mining and Applications*; Cong, G., Peng, W.C., Zhang, W.E., Li, C., Sun, A., Eds.; Springer: Cham, Switzerland, 2017; pp. 727–740. https://doi.org/10.1007/978-3-319-69179-4_51.
53. Tang, X.; Chen, H.; Xiang, W.; Yang, J.; Zou, M. Short-Term Load Forecasting Using Channel and Temporal Attention Based Temporal Convolutional Network. *Electr. Power Syst. Res.* **2022**, *205*, 107761. <https://doi.org/10.1016/j.epsr.2021.107761>.
54. Shaikh, A.K.; Nazir, A.; Khalique, N.; Shah, A.S.; Adhikari, N. A new approach to seasonal energy consumption forecasting using temporal convolutional networks. *Results Eng.* **2023**, *19*, 101296. <https://doi.org/10.1016/j.rineng.2023.101296>.
55. Ghimire, S.; Deo, R.C.; Casillas-Pérez, D.; Salcedo-Sanz, S.; Acharya, R.; Dinh, T. Electricity demand uncertainty modeling with Temporal Convolution Neural Network models. *Renew. Sustain. Energy Rev.* **2025**, *209*, 115097. <https://doi.org/10.1016/j.rser.2024.115097>.
56. UCI Machine Learning Repository. UCI Machine Learning Repository. Available online: <https://archive.ics.uci.edu/> (accessed on 9 February 2026).
57. Hua, Q.; Fan, Z.; Mu, W.; Cui, J.; Xing, R.; Liu, H.; Gao, J. A Short-Term Power Load Forecasting Method Using CNN-GRU with an Attention Mechanism. *Energies* **2024**, *18*, 106. <https://doi.org/10.3390/en18010106>.
58. Kollia, I.; Kollias, S. A Deep Learning Approach for Load Demand Forecasting of Power Systems. In *Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence (SSCI)*; IEEE: Piscataway, NJ, USA, 2018; pp. 912–919. <https://doi.org/10.1109/SSCI.2018.8628644>.
59. Nguyen, Q.V.; Fernandez, J.D.; Menci, S.P. Spatiotemporal Graph Neural Networks in short term load forecasting: Does adding Graph Structure in Consumption Data Improve Predictions? *arXiv* **2025**. <https://doi.org/10.48550/arXiv.2502.12175>.

60. Sun, C.; Ning, Y.; Shen, D.; Nie, T. Graph Neural Network-Based Short-Term Load Forecasting with Temporal Convolution. *Data Sci. Eng.* **2023**, *9*, 113–132. <https://doi.org/10.1007/s41019-023-00233-8>.
61. Chan, J.W.; Yeo, C.K. A Transformer based approach to electricity load forecasting. *Electr. J.* **2024**, *37*, 107370. <https://doi.org/10.1016/j.tej.2024.107370>.
62. Giacomazzi, E.; Haag, F.; Hopf, K. Short-Term Electricity Load Forecasting Using the Temporal Fusion Transformer: Effect of Grid Hierarchies and Data Sources. In *Proceedings of the Proceedings of the 14th ACM International Conference on Future Energy Systems*; ACM: New York, NY, USA, 2023; pp. 353–360. <https://doi.org/10.1145/3575813.3597345>.
63. Liao, H.; Wu, Z.; Ye, H.; Hu, P.; Du, W.; Tang, Y. Energy Demand Forecasting Model Based on Remember's Self-Attention Mechanism and Deep Learning. In *Proceedings of the 2024 IEEE 4th International Conference on Electronic Technology, Communication and Information (ICETCI)*; IEEE: Piscataway, NJ, USA, 2024; pp. 946–950. <https://doi.org/10.1109/ICETCI61221.2024.10594520>.
64. Zhu, L.; Gao, J.; Zhu, C.; Deng, F. Short-term power load forecasting based on spatial-temporal dynamic graph and multi-scale Transformer. *J. Comput. Des. Eng.* **2025**, *12*, 92–111. <https://doi.org/10.1093/jcde/qwaf013>.
65. Feng, D.; Li, D.; Zhou, Y.; Wang, W. MLFGCN: Short-term residential load forecasting via graph attention temporal convolution network. *Front. Neurobotics* **2024**, *18*, 1461403. <https://doi.org/10.3389/fnbot.2024.1461403>.
66. Zhou, R.; Zhang, X. Short-term power load forecasting based on ARIMA-LSTM. *J. Phys. Conf. Ser.* **2024**, *2803*, 012002. <https://doi.org/10.1088/1742-6596/2803/1/012002>.
67. Bedi, J.; Toshniwal, D. Empirical Mode Decomposition Based Deep Learning for Electricity Demand Forecasting. *IEEE Access* **2018**, *6*, 49144–49156. <https://doi.org/10.1109/ACCESS.2018.2867681>.
68. Bashir, T.; Haoyong, C.; Tahir, M.F.; Liqiang, Z. Short term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN. *Energy Rep.* **2022**, *8*, 1678–1686. <https://doi.org/10.1016/j.egy.2021.12.067>.
69. Son, N.; Shin, Y. Short- and Medium-Term Electricity Consumption Forecasting Using Prophet and GRU. *Sustainability* **2023**, *15*, 15860. <https://doi.org/10.3390/su152215860>.
70. Gupta, T.; Bhatia, R.; Sharma, S. An ensemble-based enhanced short and medium term load forecasting using optimized missing value imputation. *Sci. Rep.* **2025**, *15*, 21857. <https://doi.org/10.1038/s41598-025-06610-9>.
71. Cheng, S.; Shi, J.; Cheng, Q.; Zhou, X.; Zeng, S. Hybrid Model for Medium-Term Load Forecasting in Urban Power Grids. *Energies* **2025**, *18*, 4378. <https://doi.org/10.3390/en18164378>.
72. Yang, J.; Zhang, X.; Chen, W.; Rong, F. Prophet-CEEMDAN-ARBiLSTM-Based Model for Short-Term Load Forecasting. *Future Internet* **2024**, *16*, 192. <https://doi.org/10.3390/fi16060192>.
73. Hu, J.; Duan, P.; Cao, X.; Xue, Q.; Zhao, B.; Zhao, X.; Yuan, X.; Zhang, C. A multi-energy load forecasting method based on the Mixture-of-Experts model and dynamic multilevel attention mechanism. *Energy* **2025**, *324*, 135947. <https://doi.org/10.1016/j.energy.2025.135947>.
74. Fedus, W.; Zoph, B.; Shazeer, N. Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity. *J. Mach. Learn. Res.* **2022**, *23*, 1–39.
75. Jung, S.M.; Park, S.; Jung, S.W.; Hwang, E. Monthly Electric Load Forecasting Using Transfer Learning for Smart Cities. *Sustainability* **2020**, *12*, 6364. <https://doi.org/10.3390/su12166364>.
76. De Lange, M.; Aljundi, R.; Masana, M.; Parisot, S.; Jia, X.; Leonardis, A.; Slabaugh, G.; Tuytelaars, T. A continual learning survey: Defying forgetting in classification tasks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3366–3385.
77. Goodfellow, I.J.; Mirza, M.; Xiao, D.; Courville, A.; Bengio, Y. An empirical investigation of catastrophic forgetting in gradient-based neural networks. In *Proceedings of the International Conference on Learning Representations Workshops*, Scottsdale, Arizona, 2–4 May 2013.
78. Shafiuzzaman, M.; Safayet Islam, M.; Rubaith Bashar, T.M.; Munem, M.; Nahiduzzaman, M.; Ahsan, M.; Haider, J. Enhanced very short-term load forecasting with multi-lag feature engineering and prophet-XGBoost-CatBoost architecture. *Energy* **2025**, *335*, 137981. <https://doi.org/10.1016/j.energy.2025.137981>.
79. del Rey, S.; Cruz, L.; Franch, X.; Martínez-Fernández, S. Estimating Deep Learning energy consumption based on model architecture and training environment. *arXiv* **2025**. <https://doi.org/10.48550/arXiv.2307.05520>.
80. Benti, N.E.; Chaka, M.D.; Semie, A.G. Forecasting Renewable Energy Generation with Machine Learning and Deep Learning: Current Advances and Future Prospects. *Sustainability* **2023**, *15*, 7087. <https://doi.org/10.3390/su15097087>.
81. Ahmad, A.; Xiao, X.; Mo, H.; Dong, D. TFTformer: A novel transformer based model for short-term load forecasting. *Int. J. Electr. Power Energy Syst.* **2025**, *166*, 110549. <https://doi.org/10.1016/j.ijepes.2025.110549>.
82. Zhou, X.; Wang, J. Panel quantile regression neural network for electricity consumption forecasting in China: A new framework. *Taylor Fr.* **2021**, *16*, 420–442.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.